DISSERTATION

CHANNEL CODING FOR NETWORK COMMUNICATION: AN INFORMATION THEORETIC PERSPECTIVE

Submitted by

Zheng Wang

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2011

Doctoral Committee:

Advisor: J. Rockey Luo

Louis L. Scharf
Edwin K. P. Chong
Anton Betten

ABSTRACT

CHANNEL CODING FOR NETWORK COMMUNICATION: AN INFORMATION THEORETIC PERSPECTIVE

Channel coding helps a communication system to combat noise and interference by adding "redundancy" to the source message. Theoretical fundamentals of channel coding in point-to-point systems have been intensively studied in the research area of information theory, which was proposed by Claude Shannon in his celebrated work in 1948. A set of landmark results have been developed to characterize the performance limitations in terms of the rate and the reliability tradeoff bounds. However, unlike its success in point-to-point systems, information theory has not yielded as rich results in network communication, which has been a key research focus over the past two decades. Due to the limitations posed by some of the key assumptions in classical information theory, network information theory is far from being mature and complete. For example, the classical information theoretic model assumes that communication parameters such as the information rate should be jointly determined by all transmitters and receivers. Communication should be carried out continuously over a long time such that the overhead of communication coordination becomes negligible. The communication channel should be stationary in order for the coding scheme to transform the channel noise randomness into deterministic statistics. These assumptions are valid in a point-to-point system, but they do not permit an extensive application of channel coding in network systems because they have essentially ignored the dynamic nature of network communication. Network systems deal with bursty message transmissions between highly dynamic users. For various reasons, joint determination of key communication parameters before message transmission is often infeasible or expensive. Communication channels can often be non-stationary due to the dynamic communication interference generated by the network users. The objective of this work is to extend information theory toward network communication scenarios. We develop new channel coding results, in terms of the communication rate and error performance tradeoff, for several non-classical communication models, in which key assumptions made in classical channel coding are dropped or revised.

ACKNOWLEDGEMENTS

My deepest gratitude is to my advisor Dr. J. Rockey Luo. It has been a great honor to study under the supervision of Dr. Luo. I can never learn enough from his intelligence, creativity, diligence, as well as his high standards and rigorous attitude toward research. With great passion and patience, he has taught me how to think big and start small, to express my ideas, and to achieve great things despite all the obstacles and frustrations. He is my mentor not only in research, but also in career and life.

I would also like to thank my committee members, Dr. Louis L. Scharf, Dr. Edwin K. P. Chong and Dr. Anton Betten, for witnessing every important step of my Ph.D. study and providing valuable suggestions and guidance.

I also want to give sincere thanks to Dulanjalie Dhanapala for her precious friendship. She has shared every piece of my happiness and frustration since the beginning of my study in the US. Whenever I need help, she is always there for me without hesitation. During my Ph.D. study, I had the chance to meet many intelligent and helpful people in academia, from whom I have learned a great deal. Without mentioning their names, many thanks to all of them.

I owe everything to my parents who have been guiding me all the way here and have made me who I am today. Their unconditional love and infinite support are always inspiring me to pursue my goals. I am also thankful to my other family members and friends in China, who have also helped me to get through all the ups and downs in my life during the past several years.

Special thanks is given to my fiance Sean Zhang. His incredible love, commitment and support have lightened my life and motivated me to move forward with a warm heart. A new chapter of our life is about to begin, in which we have planted our faith deeply.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Channel Coding Basics

In a point-to-point communication system, messages are transmitted from the transmitter to the receiver through a communication channel, which introduces ambient noise and various forms of interference. The mathematical model of a point-to-point communication system was abstracted by Shannon in his celebrated work [1], summarized below.

Assume that the message to be transmitted, denoted by w, is randomly selected from a finite set $\{1, \cdots, W\}$, with equiprobability. At the transmitter, message w is mapped into a codeword consisting of N channel input symbols, with each symbol, denoted by X, chosen from a finite input alphabet $\mathcal{X}$. The mapping $\{1, \cdots, W\} \to \mathcal{X}^N$ is termed the encoding function of the system, usually characterized by a code book $C^{(N)}$. After encoding, a discrete-time memoryless channel maps each input symbol to an output symbol $Y \in \mathcal{Y}$, where $\mathcal{Y}$ is the finite output alphabet, following a conditional distribution $P_{Y|X}$. At the receiver, based on the channel output sequence, the decoder determines an estimate of the original message $\hat{w} \in \{1, \cdots, W\}$. The mapping $\mathcal{Y}^N \to \{1, \cdots, W\}$ is termed the decoding function of the system. With an abuse of the notation, we use $C^{(N)}$ to denote the channel coding scheme, which includes both the encoding and the decoding functions. Communication error probability is defined as the maximum conditional error probability over all possible messages. Namely,

$$P_e^{(N)} = \max_w Pr\{\hat{w} \neq w|w\}. \tag{1.1}$$

The information rate is defined as the number of encoded information nats normalized by the channel codeword length, given by $R = (\log W)/N$.[1]

Using this mathematical model, Shannon derived the fundamental limit of the communication channel, in terms of the maximum information rate it can support for reliable communication, known as the channel capacity or the Shannon capacity [1] [2]. For a discrete-time memoryless channel $P_{Y|X}$, the Shannon capacity is given by

$$\begin{aligned} \mathcal{C} &= \max_{P_X} I(X;Y) \\ &= \max_{P_X} \sum_{X,Y} P_X(X) P_{Y|X}(Y|X) \log \frac{P_X(X) P_{Y|X}(Y|X)}{P_X(X) \sum_X P_X(X) P_{Y|X}(Y|X)}, \end{aligned} \tag{1.2}$$

where $I(\cdot;\cdot)$ is the mutual information function and P_X is the input distribution. Let $W^{(N)}$ and $P_e^{(N)}$ be the number of messages and the error probability associated with coding scheme $C^{(N)}$ of codeword length N. For any information rate $R < \mathcal{C}$, there exists a sequence of channel coding schemes $C^{(N)}$ with $\lim_{N\to\infty} \frac{1}{N} \log(W^{(N)}) = R$, such that, as the codeword length N is taken to infinity, we have $\lim_{N\to\infty} P_e^{(N)} = 0$ [1]. Or in other words, the receiver can reliably recover the transmitted message. For any rate $R > \mathcal{C}$, on the other hand, for all coding scheme sequences with $\lim_{N\to\infty} \frac{1}{N} \log(W^{(N)}) = R$, the error probability is asymptotically bounded away from zero [1] [3] [4].

In [5], Feinstein derived a stronger version of the channel coding theorem, showing that, for information rate $R < \mathcal{C}$, error probability $P_e^{(N)}$ can be made to decrease to zero exponentially fast in N. The corresponding exponent is termed the error exponent, defined in [6] by,

$$E(R) = \lim_{N\to\infty} -\frac{\log P_e^{(N)}}{N}, \tag{1.3}$$

which is a function of R. Elias derived a lower bound on error exponent for binary symmetric channels (BSCs) in [7]. The bound was generalized to discrete-time memoryless channels by

[1] We use natural logarithm throughout this thesis.

Fano, known as the "random coding exponent" [8]. Gallager provided a simpler derivation of the random coding exponent in [4] and also tightened the bound in the low rate case, known as the "expurgated exponent". An upper bound on the error exponent was given by Shannon, Gallager, and Berlekamp [9] [10]. The upper bound for BSCs was improved by Litsyn [11]. Without complexity constraints, the maximum achievable error exponent for a discrete-time memoryless channel is called "Gallager's exponent", which equals the random coding exponent for high rates and the expurgated exponent for low rates [4].

Computational complexity is an important factor that determines the implementation feasibility of a channel coding scheme. To achieve Gallager's exponent, decoding complexity of the coding scheme needs to increase exponentially in the codeword length [6]. Such an exponential complexity is often unaffordable especially when the codeword length is large. Consequently, how to construct capacity-approaching channel codes with both good error performance and low coding complexity is of significant research and practical interests. The best known constructive error exponents are Forney's exponent and Blokh-Zyablov exponent (or Blokh-Zyablov bound), which can be achieved by one-level concatenated codes [12] and multi-level concatenated codes [13] with polynomial coding complexity, respectively. In [14], Guruswami and Indyk showed for BSCs that Forney's exponent can be arbitrarily approached with both the encoding complexity and the decoding complexity growing linearly in the codeword length.

1.2 Network Communications

Classical information theory was originally developed with a significant emphasis on point-to-point communication [1]. However, over the past two decades, research foci have been slowly shifted to network communication systems where multiple transmitters and receivers interact with each other to achieve joint or individual communication objectives. The complication of multi-user networking has given rise to a wide range of new research problems.

The study of multi-user information theory dates back to Shannon's other significant work [15], where the two-way channel model was investigated. Since then, various multi-user channel models have been proposed, e.g., multi-access channels [16] [17], interference channels [18] [19], broadcast channels [20] and relay channels [21] [22], etc.[2] Research emphasis has been put on developing coding theorems to characterize the rate and error performances of these systems. However, unlike the single-user case, rate and error tradeoff problems in multi-user information theory turned out to be highly challenging in general. A significant number of capacity and rate-error tradeoff problems remain open after decades of research efforts. Aside from the technical challenges, some of the key assumptions made in the classical information theoretic framework posed significant limitations that do not permit a full extension of information theoretic results, more specifically channel coding results, to network communication.

Classical information theoretic model assumes stationary communication channel. Information rate and codeword length should be predetermined at the transmitter. In network communication, however, non-stationary and unknown channel variation is commonly seen due to the dynamic nature of networking activities. If the transmitter has limited channel information, a conservative information rate must be chosen to guarantee the reliable message delivery. We will show in Chapter 2 that this inefficiency can be avoided using the fountain communication model [24], which essentially shifted the responsibility of information rate determination from the transmitter to the receiver. Rate and error tradeoff performance analysis, although originally developed for a classical communication model, can be effectively extended to fountain communication systems.

Classical multi-user information theory assumes that each transmitter is backlogged with an infinite reservoir of traffic. Before message transmission, transmitters and receiver should jointly determine their codebooks and information rates. The only responsibility of the receiver is to decode its message with the best effort. However, when messages are short and bursty, and require timely dissemination, joint codebook and rate determination among multiple users may be expensive or infeasible. Consequently, reliable message recovery at the receiver often implies a

[2] A detailed survey on multi-user information theory can be found in [23].

communication overhead that can be overly expensive or infeasible. In Chapter 3, we will show that channel coding theoretic results can be extended to packet-based random access communication systems where users do not jointly determine their channel codes and information rates.

In this work, we investigate several non-classical network communication scenarios and analyze the corresponding rate and error performance tradeoff under various complexity constraints. The research work is part of the general effort of bridging Information Theory with Network Theory by extending the classical frameworks.

1.3 Outlines

This dissertation is organized as follows.

In Chapter 2, we extend linear complexity concatenated coding schemes to fountain communication for a discrete-time memoryless channel. We derive the achievable error exponent of the proposed fountain codes. It is also shown that the proposed coding schemes possess some interesting and important properties in several multi-user fountain communication scenarios.

In Chapter 3, we develop channel coding theorems for distributed random multiple access communication over a discrete-time memoryless channel. Based on the channel coding approach proposed in [25], we derive the achievable rate and error tradeoff bound under the assumption of a finite codeword length. The result is further extended to random access communication over compound channels, where channel states are known neither at the transmitters nor at the receiver.

1.4 Publications

Journal Papers:

1. Z. Wang, J. Luo, ”Approaching Blokh-Zyablov Error Exponent with Linear-Time Encodable/Decodable Codes,” IEEE Communications Letters, Vol. 13, No. 6, pp. 438-440, June 2009.

2. Z. Wang, J. Luo, ”Fountain Communication using Concatenated Codes,” submitted to IEEE Trans. on Communications.

3. Z. Wang, J. Luo, ”Error Performance of Channel Coding in Random Access Communication,” submitted to IEEE Trans. on Information Theory.

4. Z. Wang, J. Luo, ”Channel Coding in Random Multiple Access Communication over Compound Channels,” submitted to IEEE Trans. on Information Theory.

Conference Papers:

1. Z. Wang, J. Luo, ”Concatenated Fountain Codes,” IEEE International Symposium on Information Theory, Seoul, Korea, June 2009.

2. Z. Wang, J. Luo, ”Achievable Error Exponent of Channel Coding in Random Access Communication,” IEEE International Symposium on Information Theory, Austin, TX, June 2010.

3. Z. Wang, J. Luo, ”Coding Theorems for Random Access Communication over Compound Channel,” IEEE International Symposium on Information Theory, Saint Petersburg, Russia, Aug. 2011.

Chapter 2

ERROR PERFORMANCE OF LINEAR-COMPLEXITY FOUNTAIN CODES

Fountain communication [24] [26] is a new communication model proposed for reliable data transmission over erasure channels. In a point-to-point fountain communication system, the transmitter maps a message into an infinite sequence of channel symbols, which experience *arbitrary* erasures during transmission. The receiver decodes the message after the number of received symbols exceeds certain threshold. With the help of randomized coding, fountain communication achieves the same rate and error performance over different channel erasure realizations corresponding to an identical number of received symbols. Under the assumption that the erasure statistics are unknown at the transmitter, communication duration in a fountain system is determined by the receiver, rather than by the transmitter.

The first realization of fountain codes was the Luby transform (LT) codes introduced by Luby [27] for erasure channels. LT codes can recover k information nats from $k+O\left(\sqrt{k}\ln^2(k/\delta)\right)$ encoded symbols with probability $1-\delta$ and a complexity of $O(k\ln(k/\delta))$, for any $\delta > 0$ [27]. Shokrollahi proposed the Raptor codes in [28] by combining appropriate LT codes with a pre-code. Raptor codes can recover k information nats from $k(1+\epsilon)$ encoded symbols at high probability with complexity $O\left(k\log(1/\epsilon)\right)$. LT codes and Raptor codes can achieve optimum rate with close to linear and linear complexity, respectively. However, under a fixed rate, error probabilities of the two coding schemes do not decrease exponentially in the number of received symbols. Generalization of Raptor codes from erasure channels to binary symmetric channels (BSCs) was investigated by Etesami and Shokrollahi in [29]. In [30], Shamai, Telatar and Verdú systematically extended fountain communication to arbitrary channels and showed that fountain capacity [30] and Shannon capacity take the same value for stationary memoryless channels. Achievability of fountain capacity was demonstrated in [30] using a random coding scheme whose error probability decreases exponentially in the number of received symbols. Unfortunately, the random coding scheme considered in [30] is impractical due to its exponential complexity.

In this chapter, we show that classical concatenated coding schemes can be extended to fountain communication over discrete-time memoryless channels to achieve a positive fountain error exponent (defined in Section 2.5) at any rate below the fountain capacity with a linear coding complexity. Achievable error exponents for one-level and multi-level concatenated fountain codes are derived. We show that these error exponents are close in value to their upper bounds, which are Forney's exponent [12] for one-level concatenation and Blokh-Zyablov exponent [13] for multi-level concatenation, respectively. We also show that concatenated fountain codes possess several interesting properties useful for network applications. More specifically, when one or more transmitters send common information to multiple receivers over discrete-time memoryless channels, concatenated fountain codes can often achieve near optimal rate and error performance simultaneously for all receivers even when the receivers have different prior knowledge about the transmitted message.

This chapter is organized as follows. In Section 2.1, we introduce the concatenated block codes proposed by Forney [12] and generalized by Blokh and Zyablov [13], which can achieve Forney's exponent for one-level concatenation and Blokh-Zyablov exponent for multi-level concatenation respectively, with polynomial coding complexity. We prove, in Section 2.2, that the encoding/decoding complexity of concatenated codes can be reduced to linear in the codeword length, while the Forney's and Blokh-Zyablov exponents are still arbitrarily approachable. In Section 2.3, we introduce the fountain communication model. In Section 2.4, we review the random fountain codes [30], which are basic components of the concatenated fountain coding schemes introduced in Section 2.5. Rate and error tradeoff performance of the linear complexity concatenated fountain codes is analyzed in Section 2.5. Special properties of the proposed concatenated fountain codes in network communication scenarios are discussed in Section 2.6. Proofs of the main theorems are given in Section 2.7.

2.1 Concatenated Block Codes

One-level concatenated codes were proposed by Forney in his doctoral thesis in 1966 [12]. The key idea of code concatenation is to break a long and powerful code into two relatively short codes, each of which can be easily encoded and decoded. For general discrete-time memoryless channels, one-level concatenated codes can achieve a positive error exponent, i.e., Forney's exponent, at any rate less than the channel capacity with coding complexity increasing polynomially in codeword length. Blokh and Zyablov generalized the concatenated codes from one-level to multi-level [13]. The corresponding achievable error exponent is improved as the concatenation level increases, and becomes the Blokh-Zyablov exponent as the concatenation level approaches infinity. Note that, in both Forney's and Blokh-Zyablov's schemes, codes are *serially* concatenated[1].

2.1.1 One-level Concatenated Block Codes

The idea of code concatenation is illustrated in Figure 1 [12]. Assume that the channel is discrete-time and memoryless. The message to be transmitted is first encoded by a block channel code, termed the *outer* code with codeword length N_o and rate r_o, and then encoded by another set

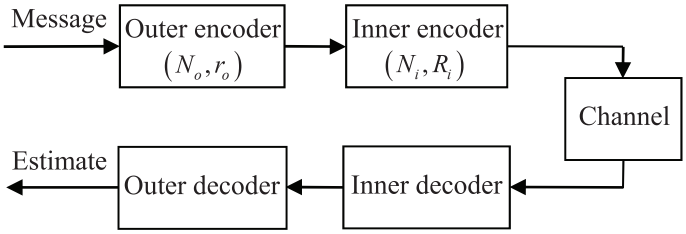
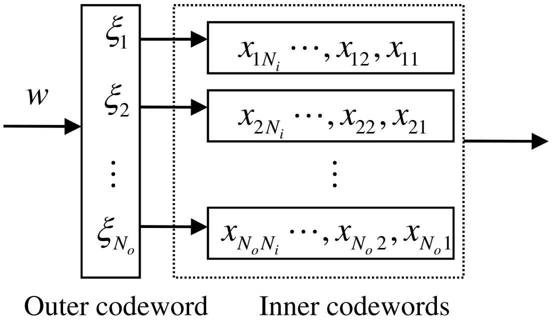


Figure 2.1: Code concatenation.

[1] Barg and Zémor proposed in [31] a *parallel* concatenated coding scheme, whose analysis is beyond the scope of this work. In the rest of the discussion, whenever concatenated coding is mentioned, we refer to Forney's and Blokh-Zyablov's serial concatenated coding schemes.

of block channel codes, termed the *inner codes*, each with codeword length N_i and rate r_i. This two-layer encoding therefore creates a *supercode* for the channel with e^{NR} codewords of length $N = N_oN_i$ and rate $R = r_oR_i$.[2] At the receiver, two decoders are concatenated to decode the inner and outer codes respectively, which eventually yields the estimate of the original message.

The encoding detail of one-level concatenated codes is shown in Figure 2.2. The encoder first maps the message w into an outer codeword with length N_o at rate r_o, denoted by $\boldsymbol{\xi} = [\xi_1, \cdots, \xi_{N_o}]$. Each outer codeword symbol ξ_k ($k \in \{1, \cdots, N_o\}$), taking $e^{N_iR_i}$ possible values, is further encoded by an inner code into N_i channel input symbols at rate R_i, denoted by $\boldsymbol{x}_k = [x_{k1}, \cdots, x_{kN_i}]$. In Forney's original scheme, Reed-Solomon code [32], a nonbinary BCH code with polynomial-time coding complexity, was chosen as the outer code. The inner codes are assumed to be the *best* block channel codes in the sense that they can achieve the optimal complexity-unconstrained error performance [12].



w
ξ_1
ξ_2
$\vdots$
ξ_{N_o}
$x_{1N_i} \cdots, x_{12}, x_{11}$
$x_{2N_i} \cdots, x_{22}, x_{21}$
$\vdots$
$x_{N_oN_i} \cdots, x_{N_o2}, x_{N_o1}$
Outer codeword
Inner codewords

Figure 2.2: One-level concatenated codes.

2.1.2 Generalized Minimum Distance Decoding

The decoding scheme used in Forney's one-level concatenation codes is the generalized minimum distance (GMD) decoding , which combines both probabilistic and algebraic decoding approaches [12]. In GMD decoding, the inner codes are first decoded using maximum likelihood

[2] To simplify the notation, we assume that e^{NR}, $e^{N_oR_o}$ and $e^{N_iR_i}$ are all integers.

decoding. The inner decoder forwards not only the maximum likelihood estimates of the outer codewords symbols, but also the corresponding reliability information (explained in detail below), to the outer decoder where multiple trials of algebraic erasures-and-errors decoding are carried out to decode the outer code. The outer decoding is a conditional searching procedure, which terminates when a certain distance criterion is met.

Assume message w is encoded by the one-level concatenated codes and transmitted over the channel. Given the channel output sequence and the channel state information, the inner decoder outputs the maximum likelihood estimate of the outer codeword, denoted by $[\hat{\xi}_1, \cdots, \hat{\xi}_{N_o}]$, along with a weight vector $[\alpha_1, \cdots, \alpha_{N_o}]$ containing the reliability information of all outer codeword symbols, with $\alpha_k \in [0, 1]$ $(k \in \{1, \cdots, N_o\})$. Let $\boldsymbol{\xi}_w = [\xi_{w1}, \cdots, \xi_{wN_o}]$ be the outer codeword corresponding to message w. Let $\boldsymbol{\alpha}$ represent the pair of the estimate vector $[\hat{\xi}_1, \cdots, \hat{\xi}_{N_o}]$ and the weight vector $[\alpha_1, \cdots, \alpha_{N_o}]$. Given $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}_w$, we define the following dot product,

$$\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w = \sum_{k=1}^{N_o} \alpha_k s(\hat{\xi}_k, \xi_{wk}), \tag{2.1}$$

where function $s(\cdot, \cdot)$ is given by

$$s(\hat{\xi}, \xi) = \begin{cases} +1, & \hat{\xi} = \xi \\ -1, & \hat{\xi} \neq \xi \end{cases}. \tag{2.2}$$

The following theorem gives the distance criterion of the GMD decoding [12].

Theorem 2.1.1. *[12, Theorem 3.1] There is at most one codeword $\boldsymbol{\xi}_w$ from a code of length N_o and rate r_o for which*

$$\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w > N_o r_o, \tag{2.3}$$

where $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w$ is defined in (2.1) and $\alpha_k \in [0, 1]$.

Note that when $\alpha_k = 1$ $(k \in \{1, \cdots, N_o\})$, the inner decoder provides no effective reliability information, but only the maximum likelihood estimates. Given the estimated outer codeword symbols, the source message can be recovered by the outer decoder using hard decision decoding, known as the *errors-only decoding* [12], with the criterion of maximizing the dot product $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w$.

If $\alpha_k \in \{0,1\}$ ($k \in \{1,\cdots,N_o\}$), the output of the inner decoder is equivalent to either an estimate (when $\alpha_k = 1$) or an erasure (when $\alpha_k = 0$). A conditional decoding scheme, called *erasures-and-errors decoding* [12], can be applied to search for the codeword that satisfies (2.3). According to Theorem 2.1.1, if such a codeword does exist, it must be unique.

In (2.3), the reliability information is further preserved by letting $\alpha_k \in [0,1]$ ($k \in \{1,\cdots,N_o\}$). The key idea of GMD decoding is to make use of the reliability information to create erasure patterns, with which multiple trials of erasures-and-errors decoding can be applied. The erasure pattern created by Forney is described as follows.

We rearrange the elements of $[\alpha_1,\cdots,\alpha_{N_o}]$ in the increasing order of their values. Let $i_1, i_2\cdots, i_{N_o}$ be the indices such that $\alpha_{i_1} \leq \alpha_{i_2} \leq \cdots \leq \alpha_{i_{N_o}}$. Define a set of N_o-dimensional vectors $\boldsymbol{q}_l = [q_l(\alpha_1),\cdots,q_l(\alpha_{N_o})]$, for $l \in \{0,1,\ldots,N_o\}$, with

$$q_l(\alpha_{i_j}) = \left\{ \begin{array}{ll} 0, & 1 \leq j \leq l \\ 1, & l+1 \leq j \leq N_o \end{array} \right. . \tag{2.4}$$

In each $\boldsymbol{q}_l$, the elements corresponding to the l least reliable symbols (with the smallest weight values) are set at 0, while the others at 1. For example, $\boldsymbol{q}_0$ is an all-ones vector, equivalent to the weight vector for the errors-only decoder; $\boldsymbol{q}_1$ erases the symbol with the smallest weight by setting the corresponding element at 0, while the others elements equals 1; $\boldsymbol{q}_2$ erases the two least reliable symbols in the same fashion, etc. The vectors $\boldsymbol{q}_l$ therefore can be regarded as the weight vectors of the erasures-and-errors decoder with various numbers of erasures. Similarly, we define the following dot product,

$$\boldsymbol{q}_l \cdot \boldsymbol{\xi}_w = \sum_{k=1}^{N_o} q_l(\alpha_k) s(\hat{\xi}_k, \xi_{wk}) = \sum_{k=1}^{N_o} q_l(\alpha_k) s_k. \tag{2.5}$$

Theorem 2.1.2. *[12, Theorem 3.2] If $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w > N_o r_o$, then, for some l, $\boldsymbol{q}_l \cdot \boldsymbol{\xi}_w > N_o r_o$.*

Theorem 2.1.1 and Theorem 2.1.2 imply that by going through up to N_o+1 rounds of erasures-and-errors decoding, each corresponding to a $\boldsymbol{q}_l$ vector, an unique codeword $\boldsymbol{\xi}_w$ satisfying (2.3) can be found, if there exists one.

2.1.3 Theoretical Performance Analysis

It was proved in [12] that, for any information rate less than Shannon capacity, denoted by $\mathcal{C}$, Forney's one-level concatenated codes can achieve a positive error exponent with encoding/decoding complexity increasing polynomially in the overall codeword length. The maximum achievable error exponent, known as Forney's exponent, is given by,

$$E_c(R) = \max_{r_o \in \left[\frac{R}{\mathcal{C}}, 1\right]} (1-r_o) E_G\left(\frac{R}{r_o}\right), \tag{2.6}$$

where $E_G\left(\frac{R}{r_o}\right)$ is the Gallager's exponent [4], defined as

$$\begin{aligned}
E_G\left(\frac{R}{r_o}\right) &= \max_{P_X} E_L\left(\frac{R}{r_o}, P_X\right), \\
E_L\left(\frac{R}{r_o}, P_X\right) &= \begin{cases} \max_{\rho\geq 1}\left\{-\rho\frac{R}{r_o} + E_x(\rho, P_X)\right\} & 0 \leq \frac{R}{r_o} < R_x \\ -\frac{R}{r_o} + E_0(1, P_X) & R_x \leq \frac{R}{r_o} < R_{\text{crit}} \\ \max_{0\leq\rho\leq 1}\left\{-\rho\frac{R}{r_o} + E_0(\rho, P_X)\right\} & R_{\text{crit}} \leq \frac{R}{r_o} < \mathcal{C} \end{cases}, \\
E_0(\rho, P_X) &= -\log\sum_Y\left(\sum_X P_X(X) P_{Y|X}(Y|X)^{\frac{1}{1+\rho}}\right)^{(1+\rho)}, \\
E_x(\rho, P_X) &= -\rho\log\sum_{X,X'} P_X(X)P_X(X') \\
&\quad \times\left(\sum_Y\sqrt{P_{Y|X}(Y|X)P_{Y|X}(Y|X')}\right)^{1/\rho}.
\end{aligned} \tag{2.7}$$

In (2.7), P_X is the input distribution of the inner codes; $P_{Y|X}$ is the channel conditional probability function; the definitions of rate thresholds R_{crit} and R_x can be found in [4]. The error exponent in (2.6) will be reduced by half if the hard decision (errors-only) decoding is used instead of the GMD decoding [12].

To investigate the coding complexity of the one-level concatenated codes, we assume that the outer codeword length N_o increases exponentially fast as the inner codeword length N_i, i.e., $N_o = O(e^{N_i})$. Since the Reed-Solomon code, used by Forney as the outer code, is a linear block code, it is straightforward to verify the polynomial overall encoding complexity. As for the decoding complexity, the inner maximum likelihood decoding complexity grows exponentially in N_i, and therefore polynomially (more specifically, quadratically) in N_o. Furthermore, as shown in Section

2.1.2, the outer erasures-and-errors decoding needs to be carried out for a number of trials upper-bounded by N_o. Each trial, for Reed-Solomon outer code, requires decoding complexity of $O(N_o^3)$ [12][3]. The overall decoding complexity is therefore $O(N_o^4)$, which is polynomial in the overall codeword length $N = N_o N_i$.

2.1.4 Multi-level Extension

For a positive integer m, an m-level concatenated code consists of m outer codes of length N_o, and N_o inner codes each with length N_i [13] [35]. As shown in Figure 2.3, the message is first encoded by m outer codes, yielding the outer codewords $\left[\xi_1^j, \cdots, \xi_{N_o}^j\right]$ for $j \in \{1, \cdots, m\}$. For $k \in \{1, \cdots, N_o\}$, the corresponding symbols of all outer codewords $\left[\xi_k^1, \cdots, \xi_k^m\right]$ are then regarded as a macro message, which is mapped into an inner codeword of N_i channel input symbols, denoted by $[x_{k1}, \cdots, x_{kN_i}]$. The overall codeword length of this m-level concatenated code is therefore $N = N_o N_i$. Let r_{oj} ($j \in \{1, \cdots, m\}$) be the rates of the outer codes and assume that all inner codes have the same rate R_i. The overall code rate is then given by $R_i \sum_{j=1}^{m} r_{oj}$.

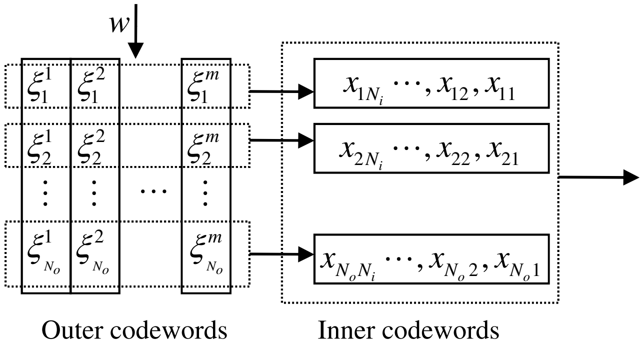

w
ξ_1^1 ξ_1^2 ξ_1^m
ξ_2^1 ξ_2^2 ξ_2^m
$\xi_{N_o}^1$ $\xi_{N_o}^2$ $\xi_{N_o}^m$
$x_{1N_i} \cdots, x_{12}, x_{11}$
$x_{2N_i} \cdots, x_{22}, x_{21}$
$x_{N_oN_i} \cdots, x_{N_o2}, x_{N_o1}$
Outer codewords
Inner codewords

Figure 2.3: m-level concatenated codes.

The decoding procedure of the m-level concatenated codes consists of m stages [35]. At each stage, GMD decoding is applied and one of the outer codes is decoded. For example, at the first stage, maximum likelihood decoding is used to decode the inner codes and to determine the

[3] Further complexity reduction on decoding the Reed-Solomon code is possible [33] [34].

estimates of all outer codeword symbols with the corresponding weight values. The information encoded by the first outer code is then decoded using multi-trial of algebraic erasures-and-errors decoding. With the assumption that the first outer code is decoded correctly, codewords corresponding to the "erroneous" messages are ruled out. At the second stage, with the "shrunk" codebooks, the information encoded by the second outer code is decoded in the same fashion, and more incorrect codewords are struck out accordingly. Such procedure continues till all outer codes are decoded.

If we assume that all outer codes have the same rate r_o, for an m-level concatenated code, the maximum achievable error exponent is given in [13] [35] by

$$E^{(m)}(R) = \max_{P_X, \frac{R}{C} \leq r_o \leq 1} \frac{\frac{R}{r_o} - R}{\frac{R}{r_o m} \sum_{i=1}^{m} \left[E_L\left(\left(\frac{i}{m}\right) \frac{R}{r_o}, P_X \right) \right]^{-1}}, \tag{2.8}$$

where $E_L(\cdot,\cdot)$ is given in (2.7). The error exponent in (2.8) becomes Forney's exponent when $m = 1$. As the concatenation level goes to infinity, this asymptotic error exponent is termed the Blokh-Zyablov exponent [13] [35], defined by,

$$E^{(\infty)}(R) = \max_{P_X, r_o \in \left[\frac{R}{C}, 1\right]} \left(\frac{R}{r_o} - R \right) \left[\int_0^{\frac{R}{r_o}} \frac{dx}{E_L(x, P_X)} \right]^{-1}. \tag{2.9}$$

Following an analysis similar to Section 2.1.3, the polynomial coding complexity of multi-level concatenated codes can be easily verified.

2.2 Error Performance of Linear-Complexity Block Codes

In this section, we illustrate the achievability of Forney's and Blokh-Zyablov exponents for general discrete-time memoryless channels with linear encoding/decoding complexity. The key result is an extension to Justesen's GMD decoding algorithm [36], which enables a low complexity integration of Guruswami-Indyk's outer code [14] into Forney's and Blokh-Zyablov's concatenated coding schemes [12] [13] reviewed in Section 2.1.

2.2.1 Linear-time Encodable/Decodable Block Codes

Guruswami and Indyk constructed in [14] a family of linear error-correction codes with coding complexity growing linearly in codeword length. These codes are near maximum distance separable (MDS) in the sense that, for any code rate $0 < r < 1$ and an arbitrarily small constant $\varepsilon > 0$, a code can be constructed to asymptotically correct up to a fraction $(1 - r - \varepsilon)/2$ of symbol errors.

By concatenating these near-MDS codes (as outer codes) with good binary inner codes, together with Justesen's GMD decoding (proposed in [36]), Forney's and Blokh-Zyablov exponents can be arbitrarily approached with linear encoding/decoding complexity for BSCs. Justesen's GMD decoding used the Hamming distance between the channel output sequence and the estimate given by the inner decoder as the reliability information. With the key assumption that the inner codeword length N_i is a constant, the outer decoder therefore only carries out a *constant* number (up to N_i) of erasures-and-errors decoding trials. According to the complexity analysis in Section 2.1.3, this is a required property for GMD decoding to achieve the overall linear decoding complexity. For BSCs, since Hamming error-correction is equivalent to (or can be transformed to an equivalent form of) maximum likelihood decoding, Forney's error exponent can be arbitrarily approached by using Guruswami-Indyk's coding scheme [14]. Note that the above results only hold for BSCs. Further revision of the GMD decoding is required in order to generalize the results to discrete-time memoryless channel, which is the main task of the next section.

2.2.2 Revised GMD Decoding

Consider Forney's one-level concatenated coding scheme over a general discrete-time memoryless channel. We use Guruswami-Indyk's linear encodable/decodable code introduced in Section 2.2.1, with length N_o and rate r_o, as the outer code. Hence, for an arbitrarily small constant $\varepsilon_1 > 0$, the outer code can correct t errors and s erasures so long as $2t + s < N_o(1 - r_o - \varepsilon_1)$. To simplify the notation, we assume that $N_o(1 - r_o - \varepsilon_1)$ is an integer. The outer code is concatenated to suitable inner codes with fixed length N_i and rate R_i. The overall codeword length and rate of the concatenated code are therefore $N = N_oN_i$ and $R = r_oR_i$, respectively.

Similar to the original GMD decoding scheme, the inner decoder outputs not only the estimate of the outer codeword $[\hat{\xi}_1, \cdots, \hat{\xi}_{N_o}]$, but also the weight vector $[\alpha_1, \cdots, \alpha_{N_o}]$. With the same definition of the dot product as in (2.1), we have the following theorem.

Theorem 2.2.1. *There is at most one codeword $\boldsymbol{\xi}_w$ that satisfies*

$$\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w > N_o(r_o + \varepsilon_1). \tag{2.10}$$

Theorem 2.2.1 is implied by Theorem 2.1.1.

Similarly, we rearrange the elements of $\boldsymbol{\alpha}$ according to their values and let $i_1, i_2, \cdots, i_{N_o}$ be the indices such that $\alpha_{i_1} \leq \alpha_{i_2} \leq \cdots \leq \alpha_{i_{N_o}}$. Let ε_2 be a positive constant with $1/\varepsilon_2$ being an integer. Define $\boldsymbol{q}_l = [q_l(\alpha_1), \ldots, q_l(\alpha_{N_o})]$, for all $l \in \{0, 1, \cdots, 1/\varepsilon_2 - 1\}$ and $j \in \{1, \cdots, N_o\}$, as

$$q_l(\alpha_{i_j}) = \begin{cases} 0 & \alpha_{i_j} \leq l\varepsilon_2 \quad \text{and} \quad i_j \leq N_o(1 - r_o - \varepsilon_1) \\ 1 & \text{otherwise} \end{cases}. \tag{2.11}$$

Recall that in the original GMD decoding scheme, vectors $\boldsymbol{q}_l$ are defined, as in (2.4), to erase symbols one by one. For the revised GMD decoding, however, we define a set of grid values $l\varepsilon_2$ ($l \in \{0, 1, \cdots, 1/\varepsilon_2\}$), and round the elements of $\boldsymbol{\alpha}$ up to the closet grid values. For each newly defined $\boldsymbol{q}_l$ in (2.11), we erase all the symbols with grid values no larger than $l\varepsilon_2$. Therefore, the number of vectors defined in (2.11) is determined by the *constant* ε_2. Note that since the outer code can correct t errors and s erasures only when $2t + s < N_o(1 - r_o - \varepsilon_1)$, there is no need to consider the $\boldsymbol{q}_l$ vectors with more than $N_o(1 - r_o - \varepsilon_1)$ zero elements.

The following theorem gives the key result that enables the revision of Forney's GMD decoder.

Theorem 2.2.2. *If $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w > N_o\left(\frac{\varepsilon_2}{2} + (r_o + \varepsilon_1)(1 - \frac{\varepsilon_2}{2})\right)$, then, for some $l \in \{0, 1, \cdots, 1/\varepsilon_2 - 1\}$, $\boldsymbol{q}_l \cdot \boldsymbol{\xi}_w > N_o(r_o + \varepsilon_1)$.*

The proof of Theorem 2.2.2 is given in Section 2.7.1.

Theorems 2.2.1 and 2.2.2 indicate that, if $\boldsymbol{\xi}_w$ is transmitted and, for some $l \in \{0, 1, \cdots, 1/\varepsilon_2 - 1\}$, it satisfies $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w > N_o\left(\frac{\varepsilon_2}{2} + (r_o + \varepsilon_1)(1 - \frac{\varepsilon_2}{2})\right)$, errors-and-erasures decoding specified by $\boldsymbol{q}_l$ (where symbols with $\boldsymbol{q}_l(\alpha_k) = 0$ are erased) will output $\boldsymbol{\xi}_w$. Since the total number of $\boldsymbol{q}_l$ vectors is upper bounded by a constant, the outer errors-and-erasures decoding only needs to be operated for a constant number of times. Consequently, a GMD decoding that carries out errors-and-erasures decoding for all $\boldsymbol{q}_l$ vectors and compares their decoding outputs can recover $\boldsymbol{\xi}_w$ with a complexity of $O(N_o)$. Since the inner code length N_i is fixed at a constant, the overall decoding complexity is $O(N)$, i.e., linear in the overall codeword length. Proving the linear encoding complexity is straightforward, because of the linear-encodable property of the outer code and the fixed inner codeword length.

The following theorem gives an error probability bound on the one-level concatenated codes with Guruswami-Indyk's outer code and the revised GMD decoding, for general discrete-time memoryless channels.

Theorem 2.2.3. *Assume inner codes achieve Gallager's error exponent in (2.7). Let vector* $\boldsymbol{\alpha}$ *be generated according to Forney's algorithm presented in [12, Section 4.2]. Let* $\boldsymbol{\xi}_w$ *e the transmitted outer codeword. For large enough* N*, error probability of the one-level concatenated codes is upper bounded by*

$$\begin{aligned} P_e &\leq Pr\left\{\boldsymbol{\alpha}\cdot\boldsymbol{\xi}_w \leq N_o\left(\frac{\varepsilon_2}{2}+(r_o+\varepsilon_1)(1-\frac{\varepsilon_2}{2})\right)\right\} \\ &\leq \exp\left[-N\left(E_c(R)-\varepsilon\right)\right]. \end{aligned} \tag{2.12}$$

where $E_c(R)$ *is Forney's error exponent given by (2.6), and* ε *is a function of* ε_1 *and* ε_2 *with* $\varepsilon \to 0$ *if* $\varepsilon_1, \varepsilon_2 \to 0$.

The proof of Theorem 2.2.3 follows an idea similar to Forney's analysis presented in [12, Section 4.2]. The decoding failure condition in [12, Section 4.2], $\boldsymbol{\alpha}\cdot\boldsymbol{\xi}_w \leq N_o r_o$ (which is supported by [12, Theorem 3.2]), should be replaced by $\boldsymbol{\alpha}\cdot\boldsymbol{\xi}_w \leq N_o\left(\frac{\varepsilon_2}{2}+(r_o+\varepsilon_1)(1-\frac{\varepsilon_2}{2})\right)$ (which is supported by Theorem 2.2.2). Since the introduced losses, ε_1 and ε_2 can be made arbitrarily small, it is straightforward to combine them into ε in (2.12), and to show that ε can also be made arbitrarily small.

The difference between Forney's and the revised GMD decoding schemes lies in the definition of errors-and-erasures decodable vectors $\boldsymbol{q}_l$, the number of which determines the decoding complexity. Forney's GMD decoding needs to carry out errors-and-erasures decoding for a number of times linear in N_o, whereas the revised GMD decoding uses a constant number. Although the idea behind the revised GMD decoding is similar to Justesen's GMD algorithm [36], Justesen's work has focused on error-correction codes where inner decoder forwards Hamming distance information (in the form of an $\boldsymbol{\alpha}$) to the outer decoder. The number of outer code decodings performed in Justesen's GMD decoding depends on the number of possible values the elements of $\boldsymbol{\alpha}$ can take, which is upper-bounded by the inner codeword length in the BSC case. However, such a bound does not hold for a general memoryless channel.

For one-level concatenated codes to approach Forney's exponent, the only requirement for the inner codes is that they should achieve Gallager's error exponent given in (2.7), for any rate below the capacity. In order to approach a better error exponent with m-level concatenated codes ($m \in \{2, 3, \cdots\}$), the inner code must possess certain special properties. Take two-level concatenated codes as an example (i.e., $m = 2$), the required property and the existence of optimal inner code are stated in the following lemma.

Lemma 2.2.4. *Consider a discrete-time memoryless channel, let $q > 0$ be an integer and P_X an input distribution. There exists a code of length N_i and rate R_i with $q^{N_i R_i}$ codewords, which are partitioned into $q^{\frac{N_i R_i}{2}}$ groups each having $q^{\frac{N_i R_i}{2}}$ codewords. Define the error probability of the code by $P_{e1}^{(2)}(R_i, P_X)$ and the maximum error probability of the codes each characterized by the codewords in a particular group of the partition by $P_{e2}^{(2)}(R_i/2, P_X)$. The error probabilities satisfy the following inequalities*

$$\lim_{N_i \to \infty} -\frac{\log P_{e1}^{(2)}(R_i, P_X)}{N_i} \geq E_L(R_i, P_X),$$
$$\lim_{N_i \to \infty} -\frac{\log P_{e2}^{(2)}(R_i/2, P_X)}{N_i} \geq E_L(R_i/2, P_X), \tag{2.13}$$

where $E_L(\cdot, \cdot)$ is given in (2.7).

Proof. We first prove the Lemma for $R_i \leq R_x$, where R_x is defined in (2.7) and in [4]. For a random block code with length N_i and $H > q^{N_i R_i}$ codewords, partition these codewords into $q^{N_i R_i/2}$ groups with at least $\lfloor H/q^{N_i R_i/2} \rfloor$ codewords in each group. Consider a particular codeword $\boldsymbol{x}_h$, $(h \in \{1, \cdots, H\})$ the following two expurgation operations [4] are performed.

In the first operation, we consider only codeword $\boldsymbol{x}_h$ and codewords that are not in the same group with $\boldsymbol{x}_h$. In other words, we temporarily strike out the codewords in the same group with $\boldsymbol{x}_h$. Define P_{eh1} as the probability of decoding error if $\boldsymbol{x}_h$ is transmitted. Let $B_1 > 0$ be a threshold such that $Pr(P_{eh1} \geq B_1) \leq 1/2$. We expurgate $\boldsymbol{x}_h$ if $P_{eh1} \geq B_1$.

Assume $\boldsymbol{x}_h$ survives the first expurgation. In the second operation, consider the codewords within the group of $\boldsymbol{x}_h$. Define by P_{eh2} the probability of decoding error if codeword $\boldsymbol{x}_h$ is transmitted. Let $B_2 > 0$ be a threshold such that $Pr(P_{eh2} \geq B_2) \leq 1/2$. We expurgate $\boldsymbol{x}_h$ if $P_{eh2} \geq B_2$.

Since

$$\begin{aligned} & Pr(P_{eh1} < B_1, P_{eh2} < B_2) \\ & \quad = Pr(P_{eh1} < B_1) Pr(P_{eh2} < B_2 | P_{eh1} < B_1) \\ & \quad = Pr(P_{eh1} < B_1) Pr(P_{eh2} < B_2) \\ & \quad \geq \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \end{aligned} \tag{2.14}$$

the probability that $\boldsymbol{x}_h$ survives two expurgation operations is at least 1/4.

With (2.14), for $R_i \leq R_x$, the conclusion of Lemma 2.2.4 follows naturally from Gallager's analysis about expurgated code in [4, Section V]. When $R_i > R_x$, at most one expurgation operation is needed. It is easily seen that the Lemma still holds. □

Lemma 2.2.4 can be extended to m-level concatenated coding schemes, with $m > 2$ as follows.

Lemma 2.2.5. *For a discrete-time memoryless channel, for any integer $m > 0$, there exists a code of rate R_i and length N_i with $q^{N_i R_i}$ codewords, where $q > 0$ is an integer. The code* *satisfies the following properties with m-level partition. Define the error probability of the code by $P_{e_1}^{(m)}(R_i, P_X)$, where P_X is a source distribution. Name the group of all codewords as the 1^{st} level group. For $1 < k \leq m$, partition the codewords in each $(k-1)^{st}$ group into $q^{\frac{N_i R_i}{m}}$ groups each having $q^{\frac{N_i R_i(m-k+1)}{m}}$ codewords. Define by $P_{e_k}^{(m)}(R_i(m-k+1)/m, P_X)$ the maximum error probability of the codes each characterized by the codewords in a particular k^{th}-level group of the partition. The error probabilities $P_{e_k}^{(m)}$ for all $1 \leq k \leq m$ satisfy*

$$\lim_{N_i \to \infty} -\frac{\log P_{e_k}^{(m)}\left(\frac{R_i(m-k+1)}{m}, P_X\right)}{N_i} \geq E_L\left(\frac{R_i(m-k+1)}{m}, P_X\right). \tag{2.15}$$

where $E_L(\cdot,\cdot)$ is given in (2.7).

The proof of Lemma 2.2.5 can be easily obtained by extending the proof of Lemma 2.2.4.

With Lemma 2.2.4 and Lemma 2.2.5, the error performance of multi-level concatenated codes, with Guruswami-Indyk outer codes and the revised GMD decoding scheme, is given in the following theorem.

Theorem 2.2.6. *For a discrete-time memoryless channel, for any $\varepsilon > 0$ and integer $m \in \{1, 2, \cdots\}$, one can construct a sequence of m-level concatenated codes whose encoding/decoding complexity is linear in N, and whose error probability is bounded by*

$$\lim_{N \to \infty} -\frac{\log P_e}{N} \geq E^{(m)}(R) - \varepsilon,$$
$$E^{(m)}(R) = \max_{P_X, r_o \in \left[\frac{R}{C}, 1\right]} \frac{\frac{R}{r_o} - R}{\frac{R}{r_o m}\sum_{i=1}^{m}\left[E_L\left(\left(\frac{i}{m}\right)\frac{R}{r_o}, P_X\right)\right]^{-1}}, \tag{2.16}$$

where $E_L(\cdot,\cdot)$ is given in (2.7).

The proof of Theorem 2.2.6 can be obtained by combining Theorem 2.2.3, Lemma 2.2.4 and the derivation of $E^{(m)}(R)$ in [13] [35].

Note that $\lim_{m\to\infty} E^{(m)}(R) = E^{(\infty)}(R)$, where $E^{(\infty)}(R)$ is the Blokh-Zyablov error exponent given in (2.9). Theorem 2.2.6 implies that, for discrete-time memoryless channels, Blokh-Zyablov error exponent can be arbitrarily approached with linear encoding/decoding complexity.

Theorem 2.2.6 holds for linear codes over BSCs [6] [37], however, the optimal inner code required by Theorem 2.2.6 is often not linear. Therefore, the concatenated codes may not be linear although their outer codes possess linearity. One may wonder what error performance can be achieved by concatenated codes if we require the inner codes should also possess a partial linear structure. The corresponding result is stated in the following theorem.

Theorem 2.2.7. *For a discrete-time memoryless channel, for any $\varepsilon > 0$ and integer $m \in \{1, 2, \cdots\}$, one can find an integer $q > 0$ and construct a sequence of m-level concatenated codes, such that each code in the sequence consists of a linear code defined on GF(q) and a mapping device that maps each GF(q) symbol to a channel symbol. The encoding/decoding complexity of the constructed codes is linear in N. The error probability is bounded by*

$$\lim_{N\to\infty} -\frac{\log P_e}{N} \geq E_r^{(m)}(R) - \varepsilon,$$

$$E_r^{(m)}(R) = \max_{P_X, r_o \in \left[\frac{R}{C}, 1\right]} \frac{\frac{R}{r_o} - R}{\frac{R}{r_o m}\sum_{i=1}^{m}\left[E_r\left((\frac{i}{m})\frac{R}{r_o}, P_X\right)\right]^{-1}} \tag{2.17}$$

where $E_r\left((\frac{i}{m})\frac{R}{r_o}, P_X\right)$ is given by

$$E_r\left(\left(\frac{i}{m}\right)\frac{R}{r_o}, P_X\right) = \max_{0\leq\rho\leq 1}\left\{-\rho\left(\frac{i}{m}\right)\frac{R}{r_o} + E_o(\rho, P_X)\right\},$$

$$E_0(\rho, P_X) = -\log\sum_{Y}\left(\sum_{X} P_X(X)P_{Y|X}(Y|X)^{\frac{1}{1+\rho}}\right)^{(1+\rho)}. \tag{2.18}$$

Proof. According to [6], one can construct a linear code on GF(q) followed by a mapping device specified in the theorem to achieve the random coding exponent. Adopting this code as the inner codes gives the desired result by following the same proof of Theorem 2.2.6. □

Theorem 2.2.7 implies that the following error exponent can be arbitrarily approached by the constructed codes with linear complexity

$$E_r^{(\infty)}(R) = \max_{P_X, r_o\in\left[\frac{R}{C}, 1\right]}\left(\frac{R}{r_o} - R\right)\left[\int_0^{\frac{R}{r_o}}\frac{dx}{E_r\left(x, P_X\right)}\right]^{-1}. \tag{2.19}$$

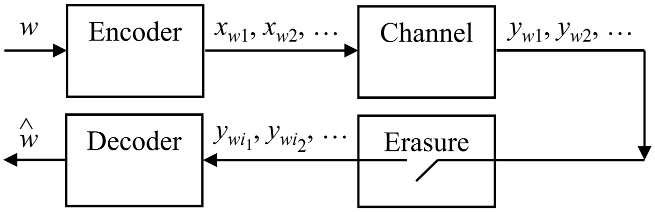


Figure 2.4: Fountain communication over a memoryless channel.

2.3 Fountain Communication Model

Consider the fountain communication system illustrated in Figure 2.4. Assume the encoder uses a fountain coding scheme [30] with W codewords to map the source message $w \in \{1, \cdots, W\}$ to an infinite channel input symbol sequence $\{x_{w1}, x_{w2}, \cdots\}$. Assume the channel is discrete-time and memoryless, characterized by the conditional point mass function (PMF) or probability density function (PDF) $P_{Y|X}$, where $X \in \mathcal{X}$ is the channel input symbol with $\mathcal{X}$ being the finite channel input alphabet, and $Y \in \mathcal{Y}$ is channel output symbol with $\mathcal{Y}$ being the finite channel output alphabet, respectively. Assume that the channel information is known at both the encoder and the decoder[4]. The channel output symbols are then passed through an erasure device which generates arbitrary erasures. Define schedule $\mathcal{N} = \{i_1, i_2, \cdots, i_{|\mathcal{N}|}\}$ as a subset of positive integers, where $|\mathcal{N}|$ is its cardinality [30]. Assume that the erasure device generates erasures only at those time indices not belonging to schedule $\mathcal{N}$. In other words, only the channel output symbols with indices in $\mathcal{N}$, denoted by $\{y_{wi_1}, y_{wi_2}, \cdots, y_{wi_{|\mathcal{N}|}}\}$, are observed by the receiver. The schedule $\mathcal{N}$ is arbitrarily chosen and unknown at the encoder.

Rate and error performance variables of the system are defined as follows. Assume that the decoder, after observing $|\mathcal{N}| \geq N$ channel symbols, outputs an estimate $\hat{w} \in \{1, 2, \cdots, W\}$ of the source message based on $\{y_{wi_1}, y_{wi_2}, \cdots, y_{wi_{|\mathcal{N}|}}\}$ and $\mathcal{N}$. We say the fountain rate of the system

[4] The case when channel information is not available at the encoder will be investigated in Section 2.6.2.

is $R = (\log W)/N$. Define error probability $P_e^{(N)}$ as,

$$P_e^{(N)} = \max_w \sup_{\mathcal{N}, |\mathcal{N}| \geq N} Pr\{\hat{w} \neq w | w, \mathcal{N}\}. \tag{2.20}$$

We say a fountain rate R is *achievable* if there exists a fountain coding scheme with $\lim_{N \to \infty} P_e^{(N)} = 0$ at rate R [30]. The exponential rate at which error probability vanishes is defined as the fountain error exponent, denoted by $E_F(R)$,

$$E_F(R) = \lim_{N \to \infty} -\frac{1}{N} \log P_e^{(N)}. \tag{2.21}$$

Define fountain capacity $\mathcal{C}_F$ as the supremum of all achievable fountain rates. It was shown in [30] that $\mathcal{C}_F$ equals Shannon capacity of the stationary memoryless channel. Note that the scaling law here is defined with respect to the number of *received* symbols.

2.4 Random Fountain Codes

In a random fountain coding scheme [30], encoder and decoder share a fountain code library $\mathcal{L} = \{C_\theta : \theta \in \Theta\}$, which is a collection of fountain codebooks C_θ indexed by a set Θ. All codebooks in the library have the same number of codewords and each codeword has an infinite number of channel input symbols. Let $C_\theta(w)_j$ be the j^{th} codeword symbol in codebook C_θ corresponding to message w, for $j \in \{1, 2, \cdots\}$. To encode the message, the encoder first selects a codebook by generating θ according to a distribution ϑ, such that the random variables $x_{w,j} : \theta \to C_\theta(w)_j$ are i.i.d. with a pre-determined input distribution P_X [30]. Then the encoder uses codebook C_θ to map the message into a codeword. We assume the actual realization of θ is known to the decoder but is unknown to the erasure device. Therefore channel erasures, although arbitrary, are independent from the codebook generation. Maximum likelihood decoding is assumed at the decoder given the knowledge of the codebook, schedule, and channel information [30]. Due to the random codebook selection, without being conditioned on θ, the error probability experienced by each message is identical. Therefore, the error probability $P_e^{(N)}$ defined in (2.20) can be written as follows [30],

$$P_e^{(N)} = \max_w \sup_{\mathcal{N}, |\mathcal{N}| \geq N} Pr\{\hat{w} \neq w | w, \mathcal{N}\} = \sup_{\mathcal{N}, |\mathcal{N}| \geq N} \frac{1}{W} \sum_w Pr\{\hat{w} \neq w | w, \mathcal{N}\}. \tag{2.22}$$

Theorem 2.4.1. *Consider fountain communication over a discrete-time memoryless channel $P_{Y|X}$. Let $\mathcal{C}_F$ be the fountain capacity. For any fountain rate $R < \mathcal{C}_F$, random fountain codes achieve the following random-coding fountain error exponent*

$$E_{Fr}(R) = \max_{P_X} E_{FL}(R, P_X), \tag{2.23}$$

where $E_{FL}(R, P_X)$ is defined as follows

$$\begin{aligned} E_{FL}(R, P_X) &= \max_{0 \le \rho \le 1} \left\{ -\rho R + E_0(\rho, P_X) \right\}, \\ E_0(\rho, P_X) &= -\log \sum_Y \left(\sum_X P_X(X) P_{Y|X}(Y|X)^{\frac{1}{1+\rho}} \right)^{(1+\rho)}. \end{aligned} \tag{2.24}$$

If the channel is continuous, then summations in (2.24) should be replaced by integrals.

Theorem 2.4.1 was claimed implicitly in, and can be shown by, the proof of [30, Theorem 2].

$E_{Fr}(R)$ given in (2.23) equals the random-coding exponent of a classical communication system over the same channel [4]. For BSCs, since random linear codes simultaneously achieve the random-coding exponent at high rates and the expurgated exponent at low rates [37], it can be easily shown that the same fountain error exponent is achievable by random linear fountain codes. However, it is not clear whether there exists an expurgation operation, such as the one proposed in [4], that is robust to the observation of any subset of channel outputs. Therefore, it is unknown whether the expurgated exponent is achievable for fountain communication over a general discrete-time memoryless channel.

2.5 Concatenated Fountain Codes

Consider a one-level concatenated fountain coding scheme illustrated in Figure 2.5. Assume that source message w can take $\lfloor \exp(NR) \rfloor$ possible values with equiprobability, where R is the targeted fountain information rate. Assume that the communication terminates after N channel output symbols are observed at the decoder. The one-level concatenated fountain code consists of an outer code and several inner codes. The encoder first encodes the message using the outer code

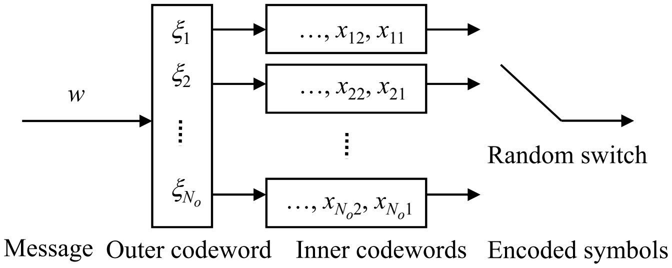


Figure 2.5: One-level concatenated fountain codes.

into an outer codeword $\{\xi_1, \xi_2, \cdots, \xi_{N_o}\}$, with N_o outer symbols, each belonging to a finite field of appropriate size. We assume that the outer code is a linear-time encodable/decodable near MDS error-correction code of rate $r_o \in (0, 1]$. The encoding and decoding complexities are linear in the number of outer codeword length N_o. An example of such linear complexity error-correction code was presented by Guruswami and Indyk in [14] and reviewed in Section 2.2.1. Each outer symbol ξ_k ($k \in \{1, \cdots, N_o\}$) can take $\left\lfloor \exp\left(\frac{N}{N_o}\frac{R}{r_o}\right)\right\rfloor$ possible values.

We use a set of random fountain codes described in Section 2.4 as the inner codes, each with $\lfloor \exp(N_i R_i) \rfloor$ codewords, where $N_i = \frac{N}{N_o}$ and $R_i = \frac{R}{r_o}$. To simplify the notations, we have assumed that N_i and N_o are both integers. We also assume that $N_o \gg N_i \gg 1$. The encoder then uses these inner codes to map each outer symbol ξ_k into an inner codeword, which is an infinite sequence of channel input symbols $\{x_{k1}, x_{k2}, \cdots\}$. The inner codewords are regarded as N_o channel input symbol queues, as shown in Figure 2.5. In each time unit, the encoder uses a random switch to pick one inner code and sends the first channel input symbol in the corresponding queue through the channel as modeled in Section 2.3. The transmitted symbol is then removed from the queue. We use θ to index the realization of the compounded randomness of codebook generation and switch selection. Let $C_\theta^{(k)}(\xi_k)_j$ be the j^{th} codeword symbol of the k^{th} inner code in codebook $\mathcal{C}_\theta^{(k)}$, corresponding to ξ_k. Let $Z_{l,\theta} \in \{1, \cdots, N_o\}$ be index of the queue that the random switch chooses at the l^{th} time unit for $l \in \{1, 2, \cdots\}$. We assume that index θ is generated according to a distribution ϑ such that random variables $x_{k,\xi_k,j} : \theta \rightarrow C_\theta^{(k)}(\xi_k)_j$ are i.i.d. with a pre-determined input distribution p_X, random variables $I_l : \theta \rightarrow Z_{l,\theta}$ are i.i.d. uniform, $x_{k,\xi_k,j}$ and

I_l are independent. The decoder is assumed to know the outer codebook and the code libraries of the inner codes. We also assume that the decoder knows the exact codebook used for each inner code and the exact order in which channel input symbols are transmitted.

Decoding starts after $N = N_o N_i$ channel output symbols are received. The decoder first distributes the received symbols to the corresponding inner codes. Assume that, for $k \in \{1, \cdots, N_o\}$, $z_k N_i$ channel output symbols are received from the k^{th} inner code, where $z_k > 0$ and $z_k N_i$ is an integer. We term z_k the "effective codeword length parameter" of the k^{th} inner code. By definition, we have $\sum_{k=1}^{N_o} z_k = N_o$. Based on z_k, and the received channel output symbols, $\{y_{ki_1}, y_{ki_2}, \ldots, y_{ki_{z_k N_i}}\}$, the decoder computes the maximum likelihood estimate $\hat{\xi}_k$ of the outer symbol ξ_k together with an optimized reliability weight $\alpha_k \in [0, 1]$. We assume that, given z_k and $\{y_{ki_1}, y_{ki_2}, \cdots, y_{ki_{z_k N_i}}\}$, reliability weight α_k is computed using Forney's algorithm presented in [12, Section 4.2]. With all the $\{\hat{\xi_k}\}$ and $\{\alpha_k\}$, the decoder then carries out the revised GMD decoding introduced in Section 2.2.2, and outputs an estimate $\hat{w}$ of the source message.

Due to the random codebook selection and the random switching, without conditioned on θ, error probabilities experienced by all messages are equal, i.e., $P_e^{(N)}$ satisfies (2.22). Compared with a classical concatenated code where all inner codes have the same length, in a concatenated fountain coding scheme, numbers of received symbols from different inner codes may be different. Consequently, error exponent achievable by one-level concatenated fountain codes, given in the following theorem, is less than Forney's exponent.

Theorem 2.5.1. *Consider fountain communication over a discrete-time memoryless channel $P_{Y|X}$ with fountain capacity $\mathcal{C}_F$. For any fountain rate $R < \mathcal{C}_F$, the following fountain error exponent can be arbitrarily approached by one-level concatenated fountain codes,*

$$E_{Fc}(R) = \max_{P_X, \frac{R}{\mathcal{C}_F} \le r_o \le 1, 0 \le \rho \le 1} (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, P_X) \left[1 - \frac{1 + r_o}{2} E_0(\rho, P_X) \right] \right), \qquad (2.25)$$

where $E_0(\rho, P_X)$ is defined in (2.24).

Encoding and decoding complexities of the one-level concatenated codes are linear in the number of transmitted symbols and the number of received symbols, respectively.

The proof of Theorem 2.5.1 is given in Section 2.7.2.

Corollary 2.5.2. *$E_{Fc}(R)$ is upper-bounded by Forney's error exponent $E_c(R)$ given in [12], and is lower-bounded by $\tilde{E}_{Fc}(R)$, defined by*

$$\tilde{E}_{Fc}(R) = \max_{P_X, \frac{R}{\mathcal{C}_F} \leq r_o \leq 1, 0 \leq \rho \leq 1} (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, P_X) \left[1 - E_0(\rho, P_X)\right] \right). \tag{2.26}$$

The bounds are asymptotically tight in the sense that

$$\lim_{R \to \mathcal{C}_F} \frac{\tilde{E}_{Fc}(R)}{E_{Fc}(R)} = 1. \tag{2.27}$$

The proof of Corollary 2.5.2 is given in Section 2.7.3.

In Figure 2.6, we illustrate $E_{Fc}(R)$, $E_c(R)$, and $\tilde{E}_{Fc}(R)$ for a BSC with crossover probability 0.1. We can see that $E_{Fc}(R)$ is closely approximated by $\tilde{E}_{Fc}(R)$, especially at rates close to the fountain capacity.

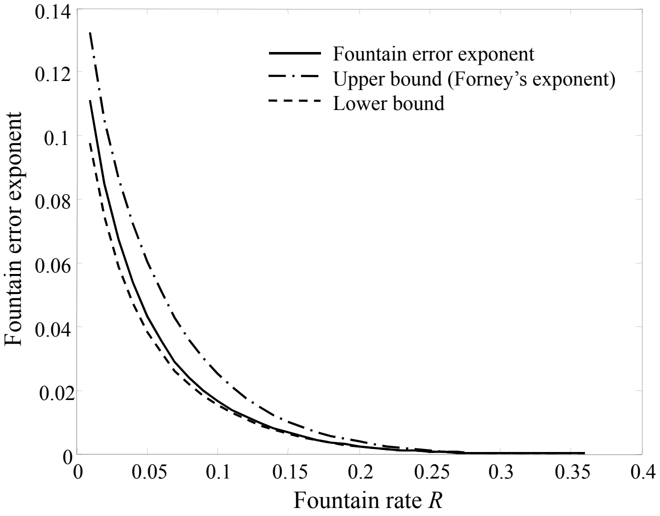


Figure 2.6: Comparison of fountain error exponent $E_{Fc}(R)$, its upper bound $E_c(R)$, and its lower bound $\tilde{E}_{Fc}(R)$.

Extending the one-level concatenated fountain codes to the multi-level concatenated fountain codes is essentially the same as in classical communication systems [13] [38] except that random fountain codes are used as inner codes in a fountain system. For a positive integer m, the achievable error exponent of an m-level concatenated fountain codes is given in the following Theorem.

Theorem 2.5.3. *Consider fountain communication over a discrete-time memoryless channel $P_{Y|X}$ with fountain capacity $\mathcal{C}_F$. For any fountain rate $R < \mathcal{C}_F$, the following fountain error exponent can be arbitrarily approached by an m-level ($m \in \{1, 2, \cdots\}$) concatenated fountain codes,*

$$\begin{aligned} E_{Fc}^{(m)}(R) &= \max_{P_X, \frac{R}{\mathcal{C}_F} \leq r_o \leq 1} \frac{\frac{R}{r_o} - R}{\frac{R}{r_o m} \sum_{i=1}^{m} \left[E_{FL}\left(\left(\frac{i}{m}\right) \frac{R}{r_o}, P_X \right) \right]^{-1}}, \\ E_{FL}\left(x, P_X\right) &= \max_{0 \leq \rho \leq 1} \left(-\rho x + E_0(\rho, P_X) \left[1 - E_0(\rho, P_X)\right] \right), \end{aligned} \tag{2.28}$$

where $E_0(\rho, P_X)$ is defined in (2.24).

For a given m, the encoding and decoding complexities of the m-level concatenated codes are linear in the number of transmitted symbols and the number of received symbols, respectively.

Theorem 2.5.3 can be proved by following the analysis of m-level concatenated codes presented in [13] [35] and replacing the error exponent of code in each concatenation level with the corresponding error exponent lower bound given in Corollary 2.5.2.

Corollary 2.5.4. *The following fountain error exponent can be arbitrarily approached by multi-level concatenated fountain codes with linear encoding/decoding complexity,*

$$E_{Fc}^{(\infty)}(R) = \max_{P_X, \frac{R}{\mathcal{C}_F} \leq r_o \leq 1} \left(\frac{R}{r_o} - R \right) \left[\int_0^{\frac{R}{r_o}} \frac{dx}{E_{FL}\left(x, P_X\right)} \right]^{-1}, \tag{2.29}$$

where $E_{FL}\left(x, P_X\right)$ is defined in (2.28).

In Figure 2.7, we illustrate $E_{Fc}^{(\infty)}(R)$ and the Blokh-Zyablov exponent $E_c^{(\infty)}(R)$ for a BSC with crossover probability 0.1. It can be seen that $E_{Fc}^{(\infty)}(R)$ does not deviate significantly from the Blokh-Zyablov exponent, which is the error exponent upper bound for multi-level concatenated fountain codes.

2.6 Network Applications

In the previous section, the rate and error performance of the concatenated codes for point-to-point communication is obtained. In this section, we extend the results to two network applications, i.e., rate compatible communication and rate combining communication.

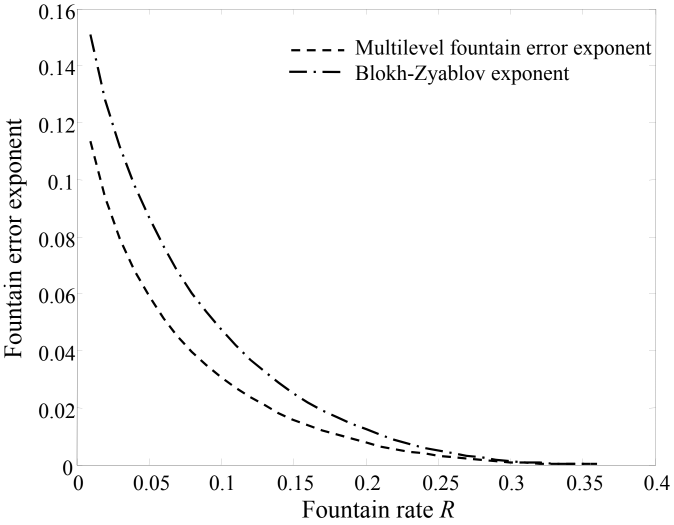


Figure 2.7: Comparison of mulit-level fountain error exponent $E_{Fc}^{(\infty)}(R)$ and the Blokh-Zyablov exponent $E_c^{(\infty)}(R)$.

2.6.1 Rate Compatible Communication

In this section, we consider the fountain communication where the receiver already has partial knowledge about the transmitted message. Take the application of software patch distribution as an example. When a significant number of patches are released, the software company may want to combine the patches together as a service pack. However, if a user already have some of the patches, he may only want to download the new patches, rather than the whole service pack. On one hand, for the convenience of the patch server, all patches of the service pack should be encoded jointly. On the other hand, for the communication efficiency of each particular user, we also want the fountain system to achieve the same rate and error performance as if only the novel part of the service pack is transmitted. We require such performance objective to be achieved simultaneously for all users, and define such a fountain communication model as the rate compatible fountain communication. We will show next that efficient rate compatible fountain communication can be achieved using a class of extended concatenated fountain codes with linear complexity.

Assume a source message w, which takes $\lfloor \exp(NR) \rfloor$ possible values, is partitioned into L sub-messages $[w_1, w_2, \cdots, w_L]$, where w_i ($i \in \{1, \cdots, L\}$) can take $\lfloor \exp(Nr_i) \rfloor$ possible values with

$\sum_i r_i = R$. Consider the following extended one-level concatenated fountain coding scheme. For each $i \in \{1, \cdots, L\}$, the encoder first uses a near MDS outer code with length N_o and rate r_o to encode sub-message w_i into an outer codeword $\{\xi_{i1}, \cdots, \xi_{iN_o}\}$, as illustrated in Figure 2.8. Next, for all $k \in \{1, \cdots, N_o\}$, the encoder combines outer codeword symbols $\{\xi_{1k}, \cdots, \xi_{Lk}\}$ into a macro symbol $\xi_k = [\xi_{1k}, \cdots, \xi_{Lk}]$. A random fountain code is then used to map ξ_k into an infinite channel input sequence $\{x_{k1}, x_{k2}, \cdots\}$.

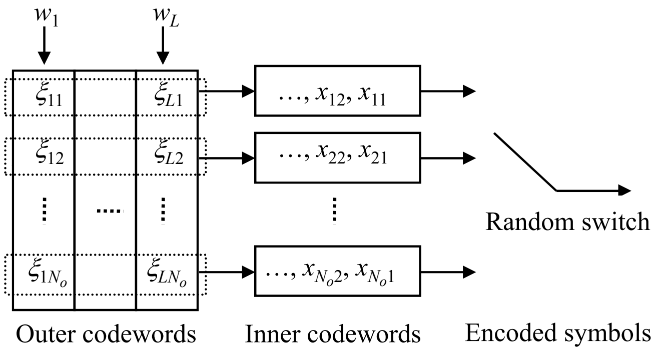


Figure 2.8: Concatenated fountain codes for rate compatible communication.

Without loss of generality, we assume that there is only one decoder (receiver) and it already has sub-messages $\{w_{l+1}, \cdots, w_L\}$, where $l \in [1, L-1]$ is an integer. The decoder estimates the source message after $N_l = N\frac{\sum_{i=1}^{l} r_i}{R}$ channel output symbols are received[5]. From the decoder's point of view, since the unknown messages $[w_1, \cdots, w_l]$ can only take $\lfloor \exp(N\sum_{i=1}^{l} r_i) \rfloor$ possible values, the *effective* fountain information rate of the system is $R_{ef} = \frac{N\sum_{i=1}^{l} r_i}{N_l} = R$. According to the known messages $\{w_{l+1}, \cdots, w_L\}$, the decoder first strikes out from fountain codebooks all codewords corresponding to the wrong messages. The extended one-level concatenated fountain code is then decoded using the same procedure as described in Section 2.5. Assume the average number of symbols received by each inner codeword $\tilde{N}_i = \frac{N_l}{N_o} = \frac{N}{N_o}\frac{\sum_{i=1}^{l} r_i}{R}$ is large enough to enable asymptotic error performance analysis. By following a similar analysis given in the proof

[5] Assume N_l and N_l/N_o are both integers.

of Theorem 2.5.1, it can be seen that error exponent $E_{Fc}(R)$ given in (2.25) can still be arbitrarily approached.

Therefore, given a rate partitioning $R = [r_1, \cdots, r_L]$, the encoder can encode the complete message irrespective of the sub-messages known at the decoder. The fountain system can achieve the same rate and error performance as if only the unknown sub-messages are encoded and transmitted. If the system has multiple receivers with different priori sub-messages, the rate and error performance tradeoff as characterized in Theorem 2.5.1 can be achieved simultaneously for all receivers. Extending this scheme to the multi-level concatenated codes is straightforward.

2.6.2 Fountain Communication over Unknown Channel

In previous sections, we have assumed that concatenated fountain codes should be optimized based on a known discrete-time memoryless channel model $P_{Y|X}$. However, such an optimization may face various challenges in practical applications. For example, suppose that a transmitter broadcasts encoded symbols to multiple receivers simultaneously. Channels experienced by different receivers may be different. Even if the channels are known, the transmitter still needs to optimize fountain codes simultaneously for multiple channels. For another example, suppose the source message (e.g., a software patch) is available at multiple servers. A user may collect encoded symbols from multiple servers separately over different channels and use these symbols to jointly decode the message. By regarding the symbols as received over a virtual channel, we want the fountain system to achieve good rate and error performance without requiring the full statistical model of the virtual channel at the transmitter. We term the communication model in the latter example the rate combining fountain communication. In both examples, the research question is whether key coding parameters can be determined without full channel knowledge at the transmitter. In this section, we show that, even when the channel state is unknown at the transmitter, it is still possible to achieve near optimal rate and error performance using concatenated fountain codes.

Consider fountain communication over a discrete-time memoryless channel $P_{Y|X}$ using one-level concatenated fountain codes. We assume the channel is symmetric, and hence the optimal input distribution P_X is known at the transmitter. Other than channel alphabets and the symmetry property, we assume channel information $P_{Y|X}$ is unknown at the transmitter, but known at the receiver. Given P_X, define $I(P_X) = I(X;Y)$ as the mutual information between the input and output of the memoryless channel. We assume the transmitter and the receiver agree on achieving a fountain information rate of $\gamma I(P_X)$ where $\gamma \in [0,1]$ is termed the normalized fountain rate, known at the transmitter.

Recall from the proof of Theorem 2.5.1 that, if $P_{Y|X}$ is known at the transmitter, the outer code rate r_o can be predetermined at the transmitter and the following error exponent can be arbitrarily approached,

$$\begin{aligned} E_{Fc}(\gamma, P_X) &= \max_{0 \le r_o \le 1} E_{Fc}(\gamma, P_X, r_o), \\ E_{Fc}(\gamma, P_X, r_o) &= \max_{0 \le \rho \le 1} (1-r_o) I(P_X) \\ &\quad \times \left(-\rho \frac{\gamma}{r_o} + \frac{E_0(\rho, P_X)}{I(P_X)} \left[1 - \frac{1+r_o}{2} E_0(\rho, P_X) \right] \right). \qquad (2.30) \end{aligned}$$

Without $P_{Y|X}$ at the transmitter, the optimal r_o cannot be derived. However, with the knowledge of γ, we can set a suboptimal outer code rate by letting $r_o = \frac{\sqrt{\gamma^2+8\gamma}-\gamma}{2}$ and define the corresponding error exponent by

$$E_{Fcs}(\gamma, P_X) = E_{Fc}\left(\gamma, P_X, r_o = \frac{\sqrt{\gamma^2+8\gamma}-\gamma}{2} \right). \qquad (2.31)$$

The following theorem indicates that $E_{Fcs}(\gamma, P_X)$ approaches $E_{Fc}(\gamma, P_X)$ asymptotically as γ approaches 1.

Theorem 2.6.1. *Given the discrete-time memoryless channel $P_{Y|X}$ and a source distribution P_X, the following limit holds,*

$$\lim_{\gamma \to 1} \frac{E_{Fcs}(\gamma, P_X)}{E_{Fc}(\gamma, P_X)} = 1. \qquad (2.32)$$

The proof of Theorem 2.6.1 is given in Section 2.7.4.

In Figure 2.9, we plot $E_{Fcs}(\gamma, P_X)$ and $E_{Fc}(\gamma, P_X)$ for BSC with crossover probability 0.1. It can be seen that setting r_o at $r_o = \frac{\sqrt{\gamma^2+8\gamma}-\gamma}{2}$ is near optimal for all normalized fountain rate values. Indeed, computer simulations suggest that such an optimality conclusion applies to a wide range of channels over a wide range of fountain rates. However, further investigation on this issue is outside the scope of our research.

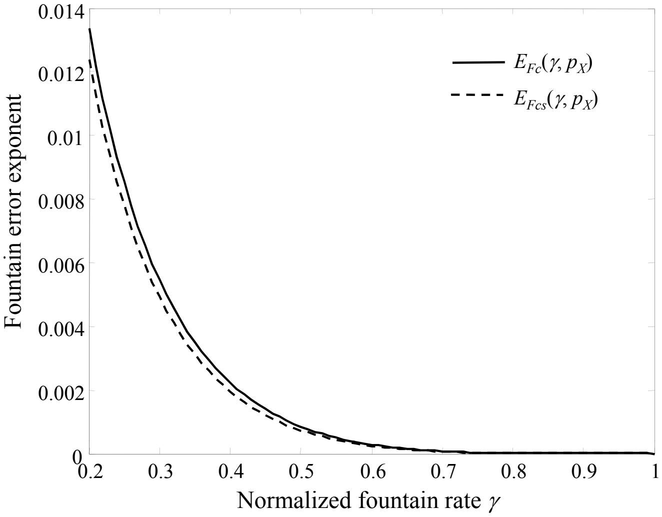


Figure 2.9: Error exponents achieved by optimal r_o and suboptimal $r_o = \frac{\sqrt{\gamma^2+8\gamma}-\gamma}{2}$ versus normalized fountain rate γ.

2.7 Proofs

In this section, the proofs of the main theorems are provided.

2.7.1 Proof of Theorem 2.2.2

Define $L = N_o(1 - r_o - \varepsilon_1)$, which is assumed to be an integer. Define a set of values $c_j = (j - 1/2)\varepsilon_2$, for $j \in \{1, \cdots, 1/\varepsilon_2\}$ and an integer $p = \lceil \alpha_{i_L}/\varepsilon_2 \rceil$, where $p \in \{1, \cdots 1/\varepsilon_2\}$.[6] Note that $1/\varepsilon_2$ is an integer.

[6] Note that the value of p cannot be 0. Because if $p = 0$, i.e., $\alpha_{i_L} = 0$, then there are at least $N_o(1 - r_o - \varepsilon_1)$ zeros in vector $\boldsymbol{\alpha}$. Consequently, $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w \le N_o(r_o + \varepsilon_1) < N_o(\frac{\varepsilon_2}{2} + (r_o + \varepsilon_1)(1 - \frac{\varepsilon_2}{2}))$, which contradicts with the assumption that $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w > N_o(\frac{\varepsilon_2}{2} + (r_o + \varepsilon_1)(1 - \frac{\varepsilon_2}{2}))$.

Let

$$\begin{aligned}
\lambda_0 &= c_1 \\
\lambda_l &= c_{l+1} - c_l, 1 \leq l \leq p-1, 1 \leq p \leq 1/\varepsilon_2 \\
\lambda_p &= \alpha_{i_{L+1}} - c_p \\
\lambda_h &= \alpha_{i_{h-p+L+1}} - \alpha_{i_{h-p+L}}, \quad p < h < p + N_o - L \\
\lambda_{p+N_o(r_o+\varepsilon_1)} &= 1 - \alpha_{i_{N_o}}.
\end{aligned} \tag{2.33}$$

We have

$$\sum_{l=0}^{j-1} \lambda_l = \begin{cases} c_j & 1 \leq j \leq p \\ \alpha_{i_{j-p+L}} & p < j \leq p + N_o - L \end{cases} \tag{2.34}$$

and

$$\sum_{l=0}^{p+N_o-L} \lambda_l = 1. \tag{2.35}$$

Define a new weight vector $\tilde{\boldsymbol{\alpha}} = [\tilde{\alpha}_1, \cdots, \tilde{\alpha}_{N_o}]$ with, for $k \in \{1, \cdots, N_o\}$,

$$\tilde{\alpha}_k = \begin{cases} \text{argmin}_{c_j, 1 \leq j \leq p} |c_j - \alpha_k| & \alpha_k \leq \alpha_{i_L} \\ \alpha_k & \alpha_k > \alpha_{i_L} \end{cases}. \tag{2.36}$$

Define $\boldsymbol{p}_l = [p_l(\alpha_1), \cdots, p_l(\alpha_{N_o})]$ with $1 \leq l \leq p + N_o - L$, such that for $0 \leq l < p$

$$\boldsymbol{p}_l = \boldsymbol{q}_l, \tag{2.37}$$

and for $p \leq l \leq p + N_o - L$

$$\boldsymbol{p}_l(\alpha_k) = \begin{cases} 0 & \alpha_k \leq \alpha_{i_{l-p+L}} \\ 1 & \alpha_k > \alpha_{i_{l-p+L}} \end{cases}. \tag{2.38}$$

Thus we have

$$\tilde{\boldsymbol{\alpha}} = \sum_{l=0}^{p+N_o-L} \lambda_l \boldsymbol{p}_l. \tag{2.39}$$

Define a set of indices

$$\mathcal{U} = \{i_1, i_2, \cdots, i_L\}. \tag{2.40}$$

According to the definition of $\tilde{\alpha}_k$, for $k \notin \mathcal{U}$, $\tilde{\alpha}_k = \alpha_k$. Hence

$$\tilde{\boldsymbol{\alpha}} \cdot \boldsymbol{\xi}_w = \boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w + \sum_{k \in \mathcal{U}} (\tilde{\alpha}_k - \alpha_k) s_k. \tag{2.41}$$

Since $|\tilde{\alpha}_k - \alpha_k| \leq \varepsilon_2/2$, and $s_k = \pm 1$, we have

$$\sum_{k \in \mathcal{U}} (\tilde{\alpha}_k - \alpha_k) \, s_k \geq -\frac{L\varepsilon_2}{2}. \tag{2.42}$$

Consequently, $\boldsymbol{\alpha} \cdot \boldsymbol{\xi}_w > N_o \left(\frac{\varepsilon_2}{2} + (r_o + \varepsilon_1)(1 - \frac{\varepsilon_2}{2}) \right) = N_o - L(1 - \frac{\varepsilon_2}{2})$ implies

$$\tilde{\boldsymbol{\alpha}} \cdot \boldsymbol{\xi}_w > N_o - L = N_o(r_o + \varepsilon_1). \tag{2.43}$$

If $\boldsymbol{p}_l \cdot \boldsymbol{\xi}_w \leq N_o - L$ for all $\boldsymbol{p}_l$ vectors, then

$$\tilde{\boldsymbol{\alpha}} \cdot \boldsymbol{\xi}_w = \sum_{l=0}^{p+N_o-L} \lambda_l \boldsymbol{p}_l \cdot \boldsymbol{\xi}_w \leq (N_o - L) \sum_{l=0}^{p+N_o-L} \lambda_l = N_o - L, \tag{2.44}$$

which contradicts to (2.43). Therefore, there must be some $\boldsymbol{p}_l$ that satisfy

$$\boldsymbol{p}_l \cdot \boldsymbol{\xi}_w > N_0 - L = N_o(r_o + \varepsilon_1). \tag{2.45}$$

Since for $l \leq p$, $\boldsymbol{p}_l$ has no more than $N_o(r_o + \varepsilon_1)$ number of 1s, which implies $\boldsymbol{p}_l \cdot \boldsymbol{\xi}_w \leq N_o(r_o + \varepsilon_1)$. Therefore, the vectors that satisfy (2.45) must exist among $\boldsymbol{p}_l$ with $1 \leq l < p$. In other words, for some l, $\boldsymbol{q}_l \cdot \boldsymbol{\xi}_w > N_o(r_o + \varepsilon_1)$.

2.7.2 Proof of Theorem 2.5.1

We first introduce the basic idea of the proof.

Assume that the decoder starts decoding after receiving $N = N_o N_i$ symbols, where N_o is the length of the outer codeword, N_i is the *expected* number of received symbols from each inner code. In the following error exponent analysis, we will obtain asymptotic results by first taking N_o to infinity and then taking N_i to infinity.

Let $\boldsymbol{z}$ be an N_o-dimensional vector whose k^{th} element z_k is the effective codeword length parameter of the k^{th} inner code, for $k \in \{1, \cdots, N_o\}$. Note that $\boldsymbol{z}$ is a random vector. Let $dz > 0$ be a small constant. We define $\{z_g | z_g = ndz, n = 0, 1, \ldots, \}$ as the set of "grid values" each can be written as an non-negative integer multiplying dz. Define a point mass function (PMF) $f_Z^{(dz)}$ as follows. We first quantize each element of $\boldsymbol{z}$, for example z_k, to the closest grid value no larger than z_k. Denote the quantized $\boldsymbol{z}$ vector by $\boldsymbol{z}^{(q)}$, whose elements are denoted by $z_i^{(q)}$ for $i \in \{1, \cdots, N_o\}$.

For any grid value z_g, we define $\mathcal{I}_{z_g} = \left\{ i \middle| z_i^{(q)} = z_g \right\}$ as the set of indices corresponding to which the elements of $\boldsymbol{z}^{(q)}$ vector equal the particular z_g. Given $\boldsymbol{z}$, the empirical PMF $f_Z^{(dz)}$ is a function defined for the grid values, with $f_Z^{(dz)}(z_g) = \frac{|\mathcal{I}_{z_g}|}{N_o}$, where $|\mathcal{I}_{z_g}|$ is the cardinality of $\mathcal{I}_{z_g}$. Since $f_Z^{(dz)}$ is induced from random vector $\boldsymbol{z}$, itself is random. Let $Pr\left\{ f_Z^{(dz)} \right\}$ denote the probability that the received effective inner codeword length parameter vector $\boldsymbol{z}$ gives a particular PMF $f_Z^{(dz)}$.

Let us now consider a decoding algorithm, called "dz-decoder", which is the same as the one introduced in Section 2.5 except that the decoder, after receiving $N_i z_k$ symbols for the k^{th} inner code (for all $k \in \{1, \cdots, N_o\}$), only uses the first $N_i z_k^{(q)}$ symbols to decode the inner code. Assume that the fountain information rate R, the outer code rate r_o, and the input distribution P_X are given. Due to symmetry, it is easy to see that, without being conditioned on random variable θ (defined in Section 2.4), different $\boldsymbol{z}$ vectors corresponding to the same $f_Z^{(dz)}$ (which is indeed induced from $\boldsymbol{z}^{(q)}$) give the same error probability performance. Let $P_e\left(f_Z^{(dz)}\right)$ be the communication error probability of the dz-decoder given $f_Z^{(dz)}$. Communication error probability P_e of the dz-decoder *without* given $f_Z^{(dz)}$ can be written as,

$$P_e = \sum_{f_Z^{(dz)}} P_e\left(f_Z^{(dz)}\right) Pr\left\{ f_Z^{(dz)} \right\}. \tag{2.46}$$

For a given $f_Z^{(dz)}$, define $E_f(f_Z^{(dz)}) = -\lim_{N_i \to \infty} \lim_{N_o \to \infty} \frac{1}{N_i N_o} \log P_e\left(f_Z^{(dz)}\right)$. Consequently, we can find a constant $K_0(N_i, N_o)$, such that the following inequality holds for all $f_Z^{(dz)}$ and all N_i, N_o,

$$\begin{aligned} & P_e\left(f_Z^{(dz)}\right) \leq K_0(N_i, N_o) \exp\left(-N_i N_o E_f\left(f_Z^{(dz)}\right)\right), \\ & \lim_{N_i \to \infty} \lim_{N_o \to \infty} \frac{\log K_0(N_i, N_o)}{N_i N_o} = 0. \end{aligned} \tag{2.47}$$

Given dz, N_i, N_o, let $K_1(N_i, N_o)$ be the total number of possible quantized $\boldsymbol{z}^{(q)}$ vectors (the quantized vector of $\boldsymbol{z}$). $K_1(N_i, N_o)$ can be upper bounded by

$$K_1(N_i, N_o) \leq 2^{N_o} \frac{\left(\left\lceil \frac{N_o}{dz} \right\rceil + N_o - 1\right)!}{\left(\left\lceil \frac{N_o}{dz} \right\rceil\right)!(N_o - 1)!}. \tag{2.48}$$

In the above bound, the term $\frac{(\lceil \frac{N_o}{dz} \rceil + N_o - 1)!}{(\lceil \frac{N_o}{dz} \rceil)!(N_o-1)!}$ represents the total number of possible outcomes of assigning $\lceil \frac{N_o}{dz} \rceil$ identical balls to N_o distinctive boxes. This is the number of possible $\boldsymbol{z}^{(q)}$ vectors we can get if the received symbols are assigned to the inner codes in groups with $N_i dz$ (assumed to be an integer) symbols per group. Let us term the assumption of assigning received symbols in groups the "symbol-grouping" assumption. To relax the symbol-grouping assumption, we note that, if the number of symbols obtained by an inner code, say the k^{th} inner code, is a little less that an integer multiplication of $N_i dz$, then the quantization value $z_k^{(q)}$ obtained without the symbol-grouping assumption can be one unit less than the corresponding value with the symbol-grouping assumption. Therefore, the total number of possible $\boldsymbol{z}^{(q)}$ vectors we can get without the symbol-grouping assumption is upper bounded by 2^{N_o} multiplying the corresponding number with the symbol-grouping assumption. Note that, given dz, the right hand side of (2.48) is not a function of N_i, and it is also an upper bound on the total number of possible $f_Z^{(dz)}$ functions.

Due to Stirling's approximation [39], (2.48) implies that $\lim_{N_o \to \infty} \frac{\log K_1(N_i, N_o)}{N_o} < \infty$, and hence

$$\lim_{N_i \to \infty} \lim_{N_o \to \infty} \frac{\log K_1(N_i, N_o)}{N_i N_o} = 0. \tag{2.49}$$

Combining (2.46), (2.47) and (2.49), the error exponent of a dz-decoder is given by

$$\begin{aligned} E_{Fc} &= -\lim_{N_i \to \infty} \lim_{N_o \to \infty} \frac{\log P_e}{N_i N_o} \\ &= \min_{f_Z^{(dz)}} \left\{ E_f \left(f_Z^{(dz)} \right) - \lim_{N_i \to \infty} \lim_{N_o \to \infty} \frac{1}{N_i N_o} \log Pr \left\{ f_Z^{(dz)} \right\} \right\}. \end{aligned} \tag{2.50}$$

The rest of the proof contains four parts.

The expression of $\lim_{N_i \to \infty} \lim_{N_o \to \infty} \frac{1}{N_i N_o} \log Pr \left\{ f_Z^{(dz)} \right\}$ is derived in Part I. In Part II, we derive the expression of $E_f \left(f_Z^{(dz)} \right)$. In Part III, we use the results of the first two parts to obtain $\lim_{dz \to 0} E_{Fc}$. Complexity and the achievable error exponent of the concatenated fountain code is obtained based on the derived results in Part IV.

Part I: Let $\boldsymbol{z}(i)$ (for all $i \in \{1, \cdots, N_o\}$) be an N_o-dimensional vector with only one non-zero element corresponding to the i^{th} received symbol. If the i^{th} received symbol belongs to the k^{th}

inner code, then we let the k^{th} element of $\boldsymbol{z}(i)$ equal 1 and let all other elements equal 0. Since the random switch (illustrated in Figure 2.2) picks inner codes uniformly, we have

$$E[\boldsymbol{z}(i)] = \frac{1}{N_o}\mathbf{1}, \quad \mathrm{cov}[\boldsymbol{z}(i)] = \frac{1}{N_o}\boldsymbol{I}_{N_o} - \frac{1}{{N_o}^2}\mathbf{1}\mathbf{1}^T, \tag{2.51}$$

where $\mathbf{1}$ is an N_o-dimensional vector with all elements being one, and $\boldsymbol{I}_{N_o}$ is the identity matrix of size N_o. According to the definitions, we have $\boldsymbol{z} = \frac{1}{N_i}\sum_{i=1}^{N_iN_o}\boldsymbol{z}(i)$. Since the total number of received symbols equal N_iN_o, we must have $\mathbf{1}^T\boldsymbol{z} = N_o$.

Let $\boldsymbol{\omega}$ be a real-valued N_o-dimensional vector whose entries satisfy $-\pi\sqrt{N_iN_o} \leq \omega_k < \pi\sqrt{N_iN_o}, \forall k \in \{1, \cdots, N_o\}$. Since $\boldsymbol{z}$ equals the normalized summation of N_iN_o independently distributed vectors $\boldsymbol{z}(i)$, the characteristic function of $\sqrt{\frac{N_i}{N_o}}(\boldsymbol{z}-\mathbf{1})$, denoted by $\varphi_Z(\boldsymbol{\omega}) = E\left[\exp\left(j\sqrt{\frac{N_i}{N_o}}\boldsymbol{\omega}^T(\boldsymbol{z}-\mathbf{1})\right)\right]$, can therefore be written as

$$\begin{aligned}
\varphi_Z(\boldsymbol{\omega}) &= E\left[\exp\left(j\sqrt{\frac{N_i}{N_o}}\boldsymbol{\omega}^T(\boldsymbol{z}-\mathbf{1})\right)\right] \\
&= \prod_{i=1}^{N_iN_o} E\left[\exp\left(j\sqrt{\frac{1}{N_iN_o}}\boldsymbol{\omega}^T(\boldsymbol{z}(i)-\frac{1}{N_o}\mathbf{1})\right)\right] \\
&= \left\{E\left[\exp\left(j\sqrt{\frac{1}{N_iN_o}}\boldsymbol{\omega}^T(\boldsymbol{z}(i)-\frac{1}{N_o}\mathbf{1})\right)\right]\right\}^{N_iN_o} \\
&= \left[1 - \frac{1}{2}\frac{\|\boldsymbol{Q}^T\boldsymbol{\omega}\|^2}{N_o^2N_i} + o\left(\frac{\|\boldsymbol{Q}^T\boldsymbol{\omega}\|^2}{N_o^2N_i}\right)\right]^{N_oN_i},
\end{aligned} \tag{2.52}$$

where in the last equality, $\boldsymbol{Q}$ is a real-valued $N_o \times (N_o-1)$-dimensional matrix satisfying $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_{N_o-1}$ and $\boldsymbol{Q}^T\mathbf{1} = \mathbf{0}$, which imply $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}_{N_o} - \frac{1}{N_o}\mathbf{1}\mathbf{1}^T$. In other words, $\|\boldsymbol{Q}^T\boldsymbol{\omega}\|^2 = \boldsymbol{\omega}^T(\boldsymbol{I}_{N_o} - \frac{1}{N_o}\mathbf{1}\mathbf{1}^T)\boldsymbol{\omega}$.

Note that, since $\boldsymbol{z}$ is discrete-valued, $\varphi_Z(\boldsymbol{\omega})$ is similar to a multi-dimensional discrete-time Fourier transform of the PMF of $\sqrt{\frac{N_i}{N_o}}(\boldsymbol{z}-\mathbf{1})$. Because $\sqrt{N_iN_o}\left[\sqrt{\frac{N_i}{N_o}}(\boldsymbol{z}-\mathbf{1})\right] = \sum_{i=1}^{N_iN_o}\boldsymbol{z}(i) - N_i\mathbf{1}$ takes integer-valued entries, the $\varphi_Z(\boldsymbol{\omega})$ function is periodic in $\boldsymbol{\omega}$ in the sense that $\varphi_Z\left(\boldsymbol{\omega} + 2\pi\sqrt{N_iN_o}\boldsymbol{e}_k\right) = \varphi_Z(\boldsymbol{\omega})$, $k \in \{1, \cdots, N_o\}$, where $\boldsymbol{e}_k$ is an N_o-dimensional vector whose k^{th} entry is one and all other entries are zeros. This is why we can focus on "frequency" vector $\boldsymbol{\omega}$ with $-\pi\sqrt{N_iN_o} \leq \omega_k < \pi\sqrt{N_iN_o}, \forall k \in \{1, \cdots, N_o\}$.

Equation (2.52) implies that

$$\lim_{N_o \to \infty} \left\{ \varphi_Z(\boldsymbol{\omega}) - \exp\left(-\frac{1}{2N_o} \boldsymbol{\omega}^T \boldsymbol{Q}\boldsymbol{Q}^T \boldsymbol{\omega} \right) \right\} = 0. \tag{2.53}$$

Therefore, with a large enough N_o and for any $\boldsymbol{z}$, the probability $Pr\{\boldsymbol{z}\}$ is upper-bounded by

$$Pr\{\boldsymbol{z}\} \leq \left(\frac{1}{2\pi\sqrt{N_i N_o}} \right)^{N_o} \left(\frac{N_o}{2\pi} \right)^{\frac{N_o - 1}{2}} \exp\left(-\frac{N_i}{2} \left[\|\boldsymbol{z} - \mathbf{1}\|^2 - dz\|\mathbf{1}\|^2 \right] \right), \tag{2.54}$$

where the constant $2\pi\sqrt{N_i N_o}$ in the denominator of the first term on the right hand side of (2.54) is due to the range of $-\pi\sqrt{N_i N_o} \leq \omega_k < \pi\sqrt{N_i N_o}, \forall k \in \{1, \cdots, N_o\}$. The constant $dz\|\mathbf{1}\|^2$ in the exponent of (2.54) is added to ensure the existence of a large enough N_o to satisfy the inequality, as implied by (2.53). Inequality (2.54) further implies that

$$\begin{aligned} Pr\{\boldsymbol{z}\} \quad \leq \quad & \left(\frac{1}{2\pi\sqrt{N_i N_o}} \right)^{N_o} \left(\frac{N_o}{2\pi} \right)^{\frac{N_o - 1}{2}} \\ & \times \exp\left(-\frac{N_i}{2} \left[\|\boldsymbol{z}^{(q)} - \mathbf{1}\|^2 - 3dz\|\mathbf{1}\|^2 \right] \right), \end{aligned} \tag{2.55}$$

where $\boldsymbol{z}^{(q)}$ is the quantized version of $\boldsymbol{z}$. Consequently, the probability of $\boldsymbol{z}^{(q)}$ is upper-bounded by

$$\begin{aligned} Pr\left\{\boldsymbol{z}^{(q)}\right\} \quad \leq \quad & \lceil N_i dz \rceil^{N_o} \left(\frac{1}{2\pi\sqrt{N_i N_o}} \right)^{N_o} \left(\frac{N_o}{2\pi} \right)^{\frac{N_o - 1}{2}} \\ & \times \exp\left(-\frac{N_i}{2} \left[\|\boldsymbol{z}^{(q)} - \mathbf{1}\|^2 - 3N_o dz \right] \right). \end{aligned} \tag{2.56}$$

The probability of any PMF $f_Z^{(dz)}$ is upper-bounded by

$$\begin{aligned} Pr\left\{f_Z^{(dz)}\right\} \quad & \leq \quad K_1(N_i, N_o) Pr\left\{\boldsymbol{z}^{(q)}\right\} \\ & \leq \quad K_1(N_i, N_o) \lceil N_i dz \rceil^{N_o} \left(\frac{1}{2\pi\sqrt{N_i N_o}} \right)^{N_o} \left(\frac{N_o}{2\pi} \right)^{\frac{N_o - 1}{2}} \\ & \qquad \times \exp\left(-\frac{N_i}{2} \left[\|\boldsymbol{z}^{(q)} - \mathbf{1}\|^2 - 3N_o dz \right] \right), \end{aligned} \tag{2.57}$$

where $K_1(N_i, N_o)$ is the total number of possible $\boldsymbol{z}^{(q)}$ vectors satisfying (2.49).

From (2.57), we can see that for all $f_Z^{(dz)}$ the following inequality holds,

$$-\lim_{N_i \to \infty} \lim_{N_o \to \infty} \frac{\log Pr\left\{f_Z^{(dz)}\right\}}{N_i N_o} \geq \frac{1}{2} \sum_{z_g} \left[(z_g - 1)^2 - 3dz \right] f_Z^{(dz)}(z_g), \tag{2.58}$$

where $f_Z^{(dz)}(z_g)$ is the value of PMF $f_Z^{(dz)}$ at z_g.

Note that, because $\mathbf{1}^T\boldsymbol{z} = N_o$, for all empirical PMFs $f_Z^{(dz)}$, we have $\sum_{z_g} z_g f_Z^{(dz)}(z_g) \in [1-dz, 1]$.

Part II: Next, we will derive the expression of $E_f\left(f_Z^{(dz)}\right)$, which is the error exponent conditioned on an empirical PMF $f_Z^{(dz)}$.

Let $\boldsymbol{z}$ be a particular N_o-dimensional effective inner codeword length parameter vector following the empirical PMF $f_Z^{(dz)}$, under a given dz. Let $P_e(\boldsymbol{z})$ be the error probability given $\boldsymbol{z}$ (or $\boldsymbol{z}^{(q)}$). Let $P_e(f_Z^{(dz)})$ be the error probability given $f_Z^{(dz)}$. From the definition of the concatenated fountain codes, we can see that the inner codes are logically equivalent, so do the codeword symbols of the near MDS outer code. In other words, error probabilities corresponding to all $\boldsymbol{z}$ vectors with the same PMF $f_Z^{(dz)}$ are equal. This consequently implies that $P_e(\boldsymbol{z}) = P_e(f_Z^{(dz)})$. Therefore, when bounding $E_f\left(f_Z^{(dz)}\right)$, instead of assuming a particular $f_Z^{(dz)}$ which corresponds to multiple $\boldsymbol{z}$ vectors, we can assume a single $\boldsymbol{z}$ vector whose corresponding empirical PMF is $f_Z^{(dz)}$.

Assume that the outer code has rate r_o, and is able to recover the source message from dN_o outer symbol erasures and tN_o outer symbol errors so long as $d + 2t \leq (1 - r_o - \zeta_0)$, where $\zeta_0 > 0$ is a constant satisfying $\lim_{N_i\to\infty}\lim_{N_o\to\infty}\zeta_0 = 0$. An example of such near MDS code was introduced in [14]. Assume that, for all k, the k^{th} outer codeword symbol is ξ_k, and the k^{th} inner code reports an estimate of the outer symbol $\hat{\xi}_k$ together with a reliability weight $\alpha_k \in [0,1]$. Applying Forney's GMD decoding to the outer code [38], the source message can be recovered if the following inequality holds [12, Theorem 3.1b],

$$\sum_{k=1}^{N_o} \alpha_k \mu_k > (r_o + \zeta_0)N_o, \tag{2.59}$$

where $\mu_k = 1$ if $\hat{\xi}_k = \xi_k$, and $\mu_k = -1$ if $\hat{\xi}_k \neq \xi_k$. Consequently, error probability conditioned on the given $\boldsymbol{z}$ vector is bounded by

$$\begin{aligned} P_e(f_Z^{(dz)}) = P_e(\boldsymbol{z}) &\leq Pr\left\{\sum_{k=1}^{N_o} \alpha_k\mu_k \leq (r_o+\zeta_0)N_o\right\} \\ &\leq \min_{s\geq 0} \frac{E\left[\exp\left(-sN_i\sum_{k=1}^{N_o}\alpha_k\mu_k\right)\right]}{\exp(-sN_i(r_o+\zeta_0)N_o)}, \end{aligned} \tag{2.60}$$

where the last inequality is due to Chernoff's bound.

Given the effective inner codeword length parameter vector $\boldsymbol{z}$, random variables $\alpha_k\mu_k$ for different inner codes are independent. Therefore, (2.60) can be further written as

$$\begin{aligned} P_e(f_Z^{(dz)}) = P_e(\boldsymbol{z}) &\leq \min_{s\geq 0} \frac{\prod_{k=1}^{N_o} E\left[\exp\left(-sN_i\alpha_k\mu_k\right)\right]}{\exp(-sN_i(r_o+\zeta_0)N_o)} \\ &= \min_{s\geq 0} \frac{\exp\left(\sum_{k=1}^{N_o}\log E\left[\exp\left(-sN_i\alpha_k\mu_k\right)\right]\right)}{\exp(-sN_i(r_o+\zeta_0)N_o)}. \end{aligned} \tag{2.61}$$

Now we will derive the expression of $\log E\left[\exp\left(-sN_i\alpha_k\mu_k\right)\right]$ for the k^{th} inner code.

Assume that the effective codeword length parameter is z_k. Given z_k, whose quantized value is $z_k^{(q)}$, depending on the received channel symbols, the decoder generates the maximum likelihood outer code estimate $\hat{\xi}_k$, and generates α_k using Forney's algorithm presented in [12, Section 4.2]. Define an adjusted error exponent function $E_z(z)$ as follows.

$$E_z(z) = \max_{0\leq\rho\leq 1} -\rho\frac{R}{r_o} + zE_0(\rho, p_X), \tag{2.62}$$

where $E_0(\rho, p_X)$ is defined in (2.24). By following Forney's error exponent analysis presented in [12, Section 4.2], we obtain

$$\begin{aligned} &-\log E\left[\exp\left(-sN_i\alpha_k\mu_k\right)\right] \\ &\quad\geq \max\left\{\min\{N_iE_z\left(z_k^{(q)}\right), N_i\left(2E_z\left(z_k^{(q)}\right)-s\right), N_is\}, 0\right\} + K_2(N_i, N_o), \end{aligned} \tag{2.63}$$

where $K_2(N_i, N_o)$ is a constant satisfying $\lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{K_2(N_i,N_o)}{N_iN_o} = 0$.

Define a function $\phi(z, s)$ as follows,

$$\phi(z,s) = \left\{\begin{array}{ll} -sr_o & z, E_z(z) < s/2 \\ 2E_z(z) - (1+r_o)s & z, s/2 \leq E_z(z) < s \\ (1-r_o)s & z, E_z(z) \geq s \end{array}\right. . \tag{2.64}$$

Substitute (2.63) into (2.61), and take N_i, N_o to infinity (which implies $\zeta_0 \to 0$), we get the following bound on the conditional error exponent $E_f\left(f_Z^{(dz)}\right)$,

$$E_f\left(f_Z^{(dz)}\right) \geq \max_{s\geq 0}\sum_{z_g}\phi(z_g, s)f_Z^{(dz)}(z_g). \tag{2.65}$$

Part III: According to (2.50), (2.58) and (2.65), we have

$$\begin{aligned} E_{Fc} &\geq \min_{f_Z^{(dz)}, \sum_{z_g} z_g f_Z^{(dz)}(z_g) \in [1-dz, 1]} \left\{ E_f(f_Z^{(dz)}) + \sum_{z_g} \frac{(z_g-1)^2}{2} f_Z^{(dz)}(z_g) \right\} - \frac{3}{2}dz \\ &\geq \min_{f_Z^{(dz)}, \sum_{z_g} z_g f_Z^{(dz)}(z_g) \in [1-dz, 1]} \max_{s \geq 0} \sum_{z_g} \left(\phi(z_g, s) + \frac{(z_g-1)^2}{2} \right) f_Z^{(dz)}(z_g) - \frac{3}{2}dz. \quad (2.66) \end{aligned}$$

Define $E_{Fc}^{(0)} = \lim_{dz \to 0} E_{Fc}$. Let f_Z be a probability density function defined for $z \in [0, \infty)$. Inequality (2.66) implies that

$$\begin{aligned} E_{Fc}^{(0)} &\geq \min_{f_Z, \int_0^\infty z f_Z(z) dz = 1} \max_{s \geq 0} \int_0^\infty \left(\phi(z, s) + \frac{(z-1)^2}{2} \right) f_Z(z) dz \\ &= \max_{s \geq 0} \min_{f_Z, \int_0^\infty z f_Z(z) dz = 1} \int_0^\infty \left(\phi(z, s) + \frac{(z-1)^2}{2} \right) f_Z(z) dz. \quad (2.67) \end{aligned}$$

Assume that f_Z^* is the density function minimizing the last term in (2.67). If we can find $0 < \lambda < 1$, and two density functions $f_Z^{(1)}$, $f_Z^{(2)}$ with $\int_0^\infty z f_Z^{(1)}(z)dz = 1$, $\int_0^\infty z f_Z^{(2)}(z)dz = 1$, such that

$$f_Z^* = \lambda f_Z^{(1)} + (1-\lambda) f_Z^{(2)}, \quad (2.68)$$

then it is easy to show that the last term in (2.67) must be minimized either by $f_Z^{(1)}$ or $f_Z^{(2)}$. Since this contradicts the assumption that f_Z^* is optimum, a nontrivial decomposition like (2.68) must not be possible. Consequently, f_Z^* can take non-zero values on at most two different z values. Therefore, we can carry out the optimization in (2.67) only over the following class of f_Z functions, characterized by two variables $0 \leq z_0 \leq 1$ and $0 \leq \gamma \leq 1$,

$$f_Z(z) = \gamma \delta(z - z_0) + (1-\gamma) \delta \left(z - \frac{1 - z_0 \gamma}{1 - \gamma} \right), \quad (2.69)$$

where $\delta()$ is the impulse function.

Let us fix γ first, and consider the following lower bound on $E_{Fc}^{(0)}(\gamma)$, which is obtained by substituting (2.69) into (2.67),

$$E_{Fc}^{(0)}(\gamma) \geq \min_{0 \leq z_0 \leq 1} \max_{s \geq 0} \gamma \phi(z_0, s) + (1-\gamma) \phi \left(\frac{1 - z_0 \gamma}{1 - \gamma}, s \right) + \frac{\gamma}{1-\gamma} \frac{(1-z_0)^2}{2}. \quad (2.70)$$

Since given z_0, $\gamma\phi(z_0, s) + (1-\gamma)\phi\left(\frac{1-z_0\gamma}{1-\gamma}, s\right)$ is a linear function of s, depending on the value of γ, the optimum s^* that maximizes the right hand side of (2.70) should satisfy either $s^* = E_z(z_0)$ or $s^* = E_z\left(\frac{1-z_0\gamma}{1-\gamma}\right)$.

When $\gamma \geq \frac{1-r_o}{2}$, we have $s^* = E_z(z_0)$. This yields

$$E_{Fc}^{(0)} \geq \min_{0\leq z_0,\gamma\leq 1}\left[\frac{\gamma}{1-\gamma}\frac{(1-z_0)^2}{2} + (1-r_o)E_z(z_0)\right]. \tag{2.71}$$

When $\gamma \leq \frac{1-r_o}{2}$, we have $s^* = E_z\left(\frac{1-z_0\gamma}{1-\gamma}\right)$, which gives

$$E_{Fc}^{(0)} \geq \min_{0\leq z_0,\gamma\leq 1}\left[2\gamma E_z(z_0) + \frac{\gamma}{1-\gamma}\frac{(1-z_0)^2}{2} + (1-r_o-2\gamma)E_z\left(\frac{1-\gamma z_0}{1-\gamma}\right)\right]. \tag{2.72}$$

By substituting $E_z(z) = \max_{0\leq\rho\leq 1}[-\rho\frac{R}{r_o} + zE_0(\rho, p_X)]$ into (2.72), we get

$$\begin{aligned} E_{Fc}^{(0)} \geq \min_{0\leq z_0,\gamma\leq 1}\max_{0\leq\rho\leq 1}\Bigg\{&(1-r_o)\left[-\rho\frac{R}{r_o} + E_0(\rho, p_X)\right] \\ &- \frac{\gamma}{1-\gamma}\left[(1+r_o)(1-z_0)E_0(\rho, p_X) - \frac{(1-z_0)^2}{2}\right]\Bigg\}. \end{aligned} \tag{2.73}$$

Note that if $(1+r_o)(1-z_0)E_0(\rho, p_X) - \frac{(1-z_0)^2}{2} < 0$, then $E_{Fc}^{(0)} \geq (1-r_o)\left[-\rho\frac{R}{r_o} + E_0(\rho, p_X)\right]$ with the right hand side of the inequality equaling Forney's exponent for given p_X and r_o. This contradicts with the fact that Forney's exponent is the maximum achievable exponent for one-level concatenated codes in a classical system [12]. Therefore, we must have $(1+r_o)(1-z_0)E_0(\rho, p_X) - \frac{(1-z_0)^2}{2} \geq 0$. Consequently, the right hand sides of both (2.71) and (2.73) are minimized at the margin of $\gamma^* = \frac{1-r_o}{2}$. This gives

$$\begin{aligned} E_{Fc}^{(0)} &\geq \min_{0\leq z_0\leq 1}\left\{(1-r_o)E_z(z_0) + \frac{1-r_o}{1+r_o}\frac{(1-z_0)^2}{2}\right\} \\ &= \min_{0\leq z_0\leq 1}\max_{0\leq\rho\leq 1}\Bigg\{(1-r_o)\left(-\rho\frac{R}{r_o} + E_0(\rho, p_X)\right) \\ &\qquad + \frac{1-r_o}{1+r_o}\frac{(1-z_0)}{2}\left[(1-z_0) - 2(1+r_o)E_0(\rho, p_X)\right]\Bigg\}. \end{aligned} \tag{2.74}$$

Note that if ρ is chosen to satisfy $(1+r_o)E_0(\rho, p_X) \geq 1$, the last term in (2.74) is minimized at $z_0^* = 0$, which gives

$$E_{Fc}^{(0)} \geq \max_{0\leq\rho\leq 1}\left\{-\rho\frac{R}{r_o}(1-r_o) + \frac{1-r_o}{1+r_o}\right\}. \tag{2.75}$$

The right hand side of (2.75) is maximized at $\rho^* = 0$. However, $\rho = 0$ implies $(1+r_o)E_0(\rho, p_X) = 0 < 1$ which contradicts the assumption $(1+r_o)E_0(\rho, p_X) \geq 1$. Therefore, we can assume that $(1+r_o)E_0(\rho, p_X) \leq 1$. Consequently, the last term in (2.74) is minimized at $z_0^* = 1-(1+r_o)E_0$. This gives

$$E_{Fc}^{(0)} \geq \max_{0\leq\rho\leq 1}(1-r_o)\left(-\rho\frac{R}{r_o} + E_0(\rho, p_X)\left[1 - \frac{1+r_o}{2}E_0(\rho, p_X)\right]\right). \tag{2.76}$$

By optimizing (2.76) over p_X and r_o, it can be seen that the error exponent given in (2.25) is achievable if we first take N_o to infinity and then take N_i to infinity.

Part IV: To achieve linear coding complexity, let us assume that N_i is fixed at a large constant while N_o is taken to infinity. According to [14], it is easy to see that the encoding complexity is linear in the number of transmitted symbols[7]. At the receiver, we keep at most $2N_i$ symbols for each inner code and drop the extra received symbols. Consequently, the effective codeword length parameter of any inner code is upper-bounded by 2. Because (2.71) and (2.73) are both minimized at $\gamma^* = \frac{1-r_o}{2}$, according to (2.69), the empirical density function $f_Z(z)$ that minimizes the error exponent bound takes the form $f_Z(z) = \frac{1-r_o}{2}\delta(z-z_0) + \frac{1+r_o}{2}\delta\left(z - \frac{2-z_0(1-r_o)}{1+r_o}\right)$, with $z_0, \frac{2-z_0(1-r_o)}{1+r_o} < 2$. Therefore, upper bounding the effective codeword length parameter by 2 does not change the error exponent result. However, with $z_k \leq 2$, $\forall k$, the decoding complexity of any inner code is upper-bounded by a constant in the order of $O(\exp(2N_i))$. According to [38], the overall decoding complexity of the concatenated code is therefore linear in N_o, and hence is linear in N. Since fixing N_i causes a reduction of $\zeta_1 > 0$ in the achievable error exponent, and ζ_1 can be made arbitrarily small as we increase N_i, we conclude that fountain error exponent $E_{Fc}(R)$ given in (2.25) can be *arbitrarily approached* by one-level concatenated fountain codes with a linear coding complexity.

[7] In other words, we assume that no encoding complexity is spent on codeword symbols that are not transmitted.

2.7.3 Proof of Corollary 2.5.2

Because $0 \le r_o \le 1$, it is easy to see $\tilde{E}_{Fc}(R) \le E_{Fc}(R) \le E_c(R)$. We will next prove $\lim_{R\to\mathcal{C}_F} \frac{\tilde{E}_{F_c}(R)}{E_{F_c}(R)} = 1$.

Define

$$g(P_X, r_o, \rho) = (1-r_o)\left(-\rho\frac{R}{r_o} + E_0(\rho, P_X)\left[1 - \frac{1+r_o}{2}E_0(\rho, P_X)\right]\right), \tag{2.77}$$

such that

$$E_{F_c}(R) = \max_{P_X, \frac{R}{\mathcal{C}_F} \le r_o \le 1, 0\le\rho\le 1} g(P_X, r_o, \rho). \tag{2.78}$$

Using Taylor's expansion to expand $g(P_X, r_o, \rho)$ at $r_o = 1$ and $\rho = 0$, we get

$$g(P_X, r_o, \rho) = \sum_{i,j} \frac{1}{(i+j)!}\beta(i,j)(r_o-1)^i\rho^j, \tag{2.79}$$

where $\beta(i,j) = \left.\frac{\partial^{(i+j)}g(P_X,r_o,\rho)}{\partial r_o^i \partial\rho^j}\right|_{r_o=1,\rho=0}$, with i and j being nonnegative integers. It can be verified that $\beta(i,j) = 0$ if $i = 0$ or $j = 0$.

We also have

$$\begin{aligned}
\beta(1,1) &= \left.\left\{\frac{R}{r_o^2} - \frac{\partial E_0(\rho,P_X)}{\partial\rho} + 2r_oE_0(\rho,P_X)\frac{\partial E_0(\rho,P_X)}{\partial\rho}\right\}\right|_{r_o=1,\rho=0} = R - \mathcal{C}_F, \\
\beta(2,1) &= -2R \ne 0, \\
\beta(1,2) &= -\left.\left\{\frac{\partial^2 E_0(\rho,P_X)}{\partial\rho^2} - 2\left(\frac{\partial E_0(\rho,P_X)}{\partial\rho}\right)^2\right\}\right|_{\rho=0} \ne 0.
\end{aligned} \tag{2.80}$$

Similarly, define

$$\tilde{g}(P_X, r_o, \rho) \quad = \quad (1-r_o)\left(-\rho\frac{R}{r_o} + E_0(\rho,P_X)\left[1 - E_0(\rho,P_X)\right]\right), \tag{2.81}$$

such that

$$\tilde{E}_{F_c}(R) = \max_{P_X, \frac{R}{\mathcal{C}_F} \le r_o \le 1, 0\le\rho\le 1} \tilde{g}(P_X, r_o, \rho). \tag{2.82}$$

Using Taylor's expansion to expand $\tilde{g}(P_X, r_o, \rho)$ at $r_o = 1$ and $\rho = 0$, we get

$$\tilde{g}(P_X, r_o, \rho) = \sum_{i,j} \frac{1}{(i+j)!}\tilde{\beta}(i,j)(r_o-1)^i\rho^j. \tag{2.83}$$

where $\tilde{\beta}(i,j) = \left.\frac{\partial^{(i+j)}\tilde{g}(P_X, r_o, \rho)}{\partial r_o^i \partial \rho^j}\right|_{r_o=1,\rho=0}$. Similarly, we have $\tilde{\beta}(i,j) = 0$ if $i = 0$ or $j = 0$ and $\tilde{\beta}(1,1) = \beta(1,1) = R - \mathcal{C}_F$, $\tilde{\beta}(2,1) = \beta(2,1) \neq 0$, $\tilde{\beta}(1,2) = \beta(1,2) \neq 0$.

By L'Hospital's rule, the following equality holds,

$$\begin{aligned} &\lim_{R\to\mathcal{C}_F} \frac{\tilde{E}_{Fc}(R)}{E_{Fc}(R)} = \\ &\lim_{R\to\mathcal{C}_F, r_o\to 1, \rho\to 0} \frac{\frac{1}{2}\tilde{\beta}(1,1)(r_o-1)\rho + \frac{1}{6}\tilde{\beta}(2,1)(r_o-1)^2\rho + \frac{1}{6}\tilde{\beta}(1,2)(r_o-1)\rho^2}{\frac{1}{2}\beta(1,1)(r_o-1)\rho + \frac{1}{6}\beta(2,1)(r_o-1)^2\rho + \frac{1}{6}\beta(1,2)(r_o-1)\rho^2} = 1. \end{aligned} \tag{2.84}$$

2.7.4 Proof of Theorem 2.6.1

Define

$$\begin{aligned} &\hat{g}(\gamma, r_o, \rho) = (1-r_o)\left(\rho I(P_X)\left(1 - \frac{\gamma}{r_o}\right) + \frac{\rho^2}{2}\left(\left.\frac{\partial^2 E_0(\rho, P_X)}{\partial \rho^2}\right|_{\rho=0} - 2I^2(P_X)\right)\right), \\ &\hat{E}_{Fc}(\gamma, P_X, r_o) = \max_{0\le\rho\le 1} \hat{g}(\gamma, r_o, \rho), \\ &\hat{E}_{Fc}(\gamma, P_X) = \max_{0\le r_o\le 1} \hat{E}_{Fc}(\gamma, P_X, r_o). \end{aligned} \tag{2.85}$$

We will first prove that

$$\lim_{\gamma\to 1} \frac{E_{Fcs}(\gamma, P_X)}{\hat{E}_{Fc}(\gamma, P_X)} = 1. \tag{2.86}$$

Note that $\hat{g}(\gamma, r_o, \rho)$ is maximized at $\rho = \rho^*$, with

$$\rho^* = \frac{I(P_X)\left(1 - \frac{\gamma}{r_o}\right)}{-\left.\frac{\partial^2 E_0(\rho, P_X)}{\partial \rho^2}\right|_{\rho=0} + 2I^2(P_X)}, \tag{2.87}$$

where we have assumed $0 \le \rho^* \le 1$. This assumption is valid when r_o is also optimized. Consequently, $\hat{E}_{Fc}(\gamma, P_X, r_o)$ is maximized at $r_o = r_o^*$, with

$$r_o^* = \arg\max_{0\le r_o\le 1}(1-r_o)\left(1 - \frac{\gamma}{r_o}\right)^2 = \frac{\sqrt{\gamma^2 + 8\gamma} - \gamma}{2}. \tag{2.88}$$

Therefore,

$$\lim_{\gamma\to 1} \frac{E_{Fcs}(\gamma, P_X)}{\hat{E}_{Fc}(\gamma, P_X)} \ge \lim_{\gamma\to 1} \left[\left.\frac{E_{Fcs}(\gamma, P_X, \rho)}{\hat{g}(\gamma, P_X, \rho, r_o)}\right|_{\rho=\rho^*, r_o=r_o^*}\right] = 1. \tag{2.89}$$

Following a similar idea as the proof of Corollary 2.5.2, it can be shown that

$$\lim_{\gamma\to 1} \frac{\hat{E}_{Fc}(\gamma, P_X)}{E_{Fc}(\gamma, P_X)} = 1. \tag{2.90}$$

Combining (2.89) and (2.90), we get

$$\lim_{\gamma \to 1} \frac{E_{Fcs}(\gamma, P_X)}{E_{Fc}(\gamma, P_X)} = \lim_{\gamma \to 1} \frac{E_{Fcs}(\gamma, P_X)}{\hat{E}_{Fc}(\gamma, P_X)} \lim_{\gamma \to 1} \frac{\hat{E}_{Fc}(\gamma, P_X)}{E_{Fc}(\gamma, P_X)} \geq 1. \tag{2.91}$$

Because $E_{Fcs}(\gamma, P_X) \leq E_{Fc}(\gamma, P_X)$, (2.91) implies $\lim_{\gamma \to 1} \frac{E_{Fcs}(\gamma, P_X)}{E_{Fc}(\gamma, P_X)} = 1$.

Chapter 3

CODING THEOREMS FOR RANDOM ACCESS COMMUNICATION

In multiple access communication, two or more users (transmitters) send messages to a common receiver. The transmitted messages confront distortion both from channel noise and from multi-user interference. Two related communication models, the multi-user information theoretic model and the random access model, have been intensively studied in the literature [40].

The information theoretic multiple access model, on one hand, assumes that each user is backlogged with an infinite reservoir of traffic. Users should first jointly determine their codebooks and information rates, and then send the encoded messages to the receiver continuously over a long communication duration. The only responsibility of the receiver is to decode the messages with its best effort. Under these assumptions, channel capacity and coding theorems are proved by taking the codeword length to infinity [1] [2]. Rate and error performance tradeoffs of single user and multiple access systems were analyzed in [4] [40]. The information theoretic model uses symbol-based statistics to characterize the communication channel. Such a physical layer channel model enables rigorous understandings on the fundamental impact of channel noise and multi-user interference. However, classical coding results have been derived under the assumption of coordinated communication, in the sense of joint codebook and information rate determination among the multiple users and the receiver. Such an assumption precludes the common scenarios of short messages and bursty traffic arrivals, since in these cases the overhead of full communication coordination is often expensive or infeasible.

The random multiple access model, on the other hand, assumes bursty message arrivals. According to message availability, users independently encode their messages into packets and randomly send these packets to the receiver. It is often assumed that the transmitted packets should be correctly received if the power of the multi-user interference is below a threshold. Otherwise the receiver should report a packet collision and the involved packets are erased [41] [42]. Standard networking regards packet as the basic communication unit, and counts system throughput in packets per time slot as opposed to bits/nats per symbol. Communication channel is characterized using packet-based models, such as the collision channel model [43] and the multipacket-reception channel model [44] [45]. Although packet-based models are convenient for upper layer networking [46], their abstract forms do not permit an insightful understanding about the impact of physical layer communication to upper layer networking.

In [25], a new channel coding approach was proposed for time-slotted random multiple access communication over a discrete-time memoryless channel using a symbol-based physical layer channel model. Assume that in each time slot, each user independently encodes an arbitrary number of data units into a packet and transmits the packet to the receiver. Define the normalized number of data units per symbol as the communication rate of a user in a time slot, which is shared neither among the users nor with the receiver. It was shown in [25] that, fundamental performance limitation of the random multiple access system can be characterized using an achievable rate region in the following sense. As the codeword length goes to infinity, if the random communication rate vector of the users happens to be inside the rate region, the receiver can decode all messages with zero asymptotic error probability; if the random communication rate vector happens to be outside the rate region, the receiver can detect a packet collision with an asymptotic probability of one. The achievable rate region was shown to equal Shannon's information rate region, possibly without a convex hull operation.

In this chapter, we derive stronger versions of the coding theorems given in [25] by characterizing the achievable rate and error performance of random multiple access communication over a discrete-time memoryless channel with a *finite* codeword length. We assume that the channel state information is known at the receiver. Our work is motivated by the existing non-asymptotic channel coding results, surveyed in [47], for classical single-user communication. Following the framework of [25], we assume that the random multiple access system predetermines an "operation region" of the rate vectors in the following sense. For all communication rate vectors within the region, the system *intends* to decode the messages; while for all communication rate vectors outside the region, the system *intends* to report a packet collision. Given the operation region, there are two types of error events. If the communication rate vector is within the region, the event that the receiver fails to decode the messages correctly is defined as a decoding error event. If the communication rate vector is outside the region, the event that the receiver fails to report a collision is defined as a collision miss detection event. An achievable bound on the system error probability, defined as the maximum of the decoding error probability and the collision miss detection probability, is obtained under the assumption of a finite codeword length. We show that, if the operation region is strictly contained in an achievable rate region, then the system error probability can decrease exponentially in the codeword length. The corresponding exponent is defined as the system error exponent, whose achievable bound is obtained from the error probability bound by taking the codeword length to infinity.

Furthermore, we relax the assumption that the channel state information is perfectly known at the receiver. Since random access communication deals with bursty short messages, transmission activities of a user are often fractional. Without frequent data support, accurate real-time channel estimation and tracking become difficult at the receiver. Understanding the system performance limitation without channel state information therefore becomes essential [48]. We illustrate how previously derived coding theorems can be extended to random multiple access communication over a compound discrete-time memoryless channel [49] [50], consisting of a family (set) of channels over which the communication could take place. Both the transmitters and the receiver know about the compound channel set, but neither knows about the actual channel realization. The compound channel communication problem investigated here is different from a conventional one in the following two key aspects. First, in a conventional system, information rates are jointly determined by the transmitters and the receiver [2], while communication rates in a random access system are determined distributively and the rate information is unknown at the receiver [25]. Second, in a conventional system, in order to achieve reliable communication, the transmitted rate vector should be supported by all channel realizations in the compound set [51] [48]. In random access communication, however, even though the receiver needs to guarantee the reliability of its decoding output, the receiver also has the additional choice of reporting a collision to avoid confusing the upper layer networking [42]. This therefore allows the transmitted rate vector to be supported only by a subset of channel realizations. If the actual channel realization belongs to this subset, the receiver should decode the messages. Otherwise, the receiver should report a collision. Clearly, the decoding and collision report decisions made at the receiver are affected jointly by the communication rates of the users and the actual channel realization. The system error probability bound of such system is derived for a finite codeword length. We also show how the compound channel results help in obtaining error performance bounds for the random multiple access system where the receiver is only interested in recovering messages from a user *subset* [25]. This is based on the observation that, conditioned on the receiver not decoding messages for the rest of the users, the impact of their communication activities on the user subset of interest is equivalent to that of a compound channel.

The chapter is organized as follows. In Section 3.1, we study the coding theorems for single-user random access communication. The results are generalized to multi-user random access communications in Section 3.2. The error performance with generalized random coding scheme is analyzed in Section 3.3. In Section 3.4, we investigate the random access communication over compound channels. The corresponding results are used in Section 3.5 to study the random access system where the receiver only decodes for a subset of users. The proofs of the main theorems are given in Section 3.6.

3.1 Single-user Random Access System

For easy understanding, we will first consider single-user random access communication over a discrete-time memoryless channel. The channel is modeled by a conditional distribution function $P_{Y|X}$. Assume that time is partitioned into slots each equaling N symbol durations, which is also the length of a packet. We focus on coding within a time slot or a packet.

Suppose that the transmitter has no channel information except the channel alphabets[1]. At the beginning of each time slot, according to message availability and the MAC layer protocol, the transmitter chooses a communication rate $r \in \{r_1, \cdots, r_M\}$, in nats per symbol, without sharing this rate information with the receiver. Here $\{r_1, \cdots, r_M\}$ is a pre-determined set of rates with cardinality M, known by both the transmitter and the receiver. The transmitter then encodes $\lfloor Nr \rfloor$ data nats, denoted by a message w, into a codeword using a "random coding scheme" described as follows [25][2]. Let $\mathcal{L} = \{C_\theta : \theta \in \Theta\}$ be a library of codebooks indexed by a set Θ. Each codebook contains M classes of codewords. The i^{th} ($i \in \{1, \cdots, M\}$) codeword class contains $\lfloor e^{Nr_i} \rfloor$ codewords, each of N symbol length. Let $C_\theta(w, r)_j$ be the j^{th} codeword symbol of message and communication rate pair (w, r) in codebook C_θ, for $j \in \{1, \cdots, N\}$. The transmitter first randomly generates θ according to a distribution γ, such that random variables $X_{(w,r),j} : \theta \to C_\theta(w, r)_j$ are independently distributed according to an input distribution $P_{X|r}$[3]. The random access codebook C_θ is then used to map the message into a codeword. This is equivalent to mapping a message and rate pair (w, r) into a codeword, denoted by $\boldsymbol{x}_{(w,r)}$, of N channel input symbols. We denote $(\mathcal{L}, \gamma)$ as this random coding scheme.

We assume the receiver knows the channel $P_{Y|X}$ and the randomly generated codebook C_θ[4]. Based on this information, the receiver chooses a rate subset $\mathcal{R} \subseteq \{r_1, \cdots, r_M\}$. According to

[1] The significance of this assumption will become clear when we investigate multi-user systems.

[2] Note that the coding scheme is an extended version of the random coding introduced in [30] and Section 2.4.

[3] We allow the input distribution to be a function of communication rate. In other words, codewords corresponding to different communication rates may be generated according to different input distributions.

[4] This can be realized by sharing the codebook generation algorithm with the receiver.

the channel output symbol vector $\boldsymbol{y}$, the receiver outputs an estimated message and rate pair $(\hat{w}, \hat{r})$ if and only if $\hat{r} \in \mathcal{R}$ and a predetermined decoding error probability requirement is satisfied. Otherwise the receiver outputs a collision. Note that the term "collision" here is used to maintain consistency with the networking terminology. Throughout the paper, collision means outage, irrespective whether it is caused by multi-user interference or by excessive channel noise.

Since the receiver *intends* to decode all messages with $r \in \mathcal{R}$ and to report collision for messages with $r \notin \mathcal{R}$, we say $\mathcal{R} \subseteq \{r_1, \cdots, r_M\}$ is the "operation region" of the system. Conditioned on (w, r) is transmitted, for $r \in \mathcal{R}$, we define the decoding error probability with codeword length N as

$$P_e^{(N)}(w, r) = Pr\{(\hat{w}, \hat{r}) \neq (w, r) | (w, r)\}, \qquad \forall (w, r), r \in \mathcal{R}. \tag{3.1}$$

For $r \notin \mathcal{R}$, we define the collision miss detection probability with codeword length N as

$$\bar{P}_c^{(N)}(w, r) = 1 - Pr\{\text{“collision”} | (w, r)\}, \qquad \forall (w, r), r \notin \mathcal{R}. \tag{3.2}$$

As defined in [25], a rate region $\mathcal{R}$ is said to be *achievable* if there exists a series of random coding schemes $(\mathcal{L}^{(N)}, \gamma^{(N)})$ with the decoding error probability and the collision miss detection probability given in (3.1) and (3.2), such that

$$\begin{aligned} &\lim_{N \to \infty} P_e^{(N)}(w, r) = 0, \quad \forall (w, r), r \in \mathcal{R}, \\ &\lim_{N \to \infty} \bar{P}_c^{(N)}(w, r) = 0, \quad \forall (w, r), r \notin \mathcal{R}. \end{aligned} \tag{3.3}$$

In other words, asymptotically, the receiver can reliably decode the message if the random communication rate r is inside the rate region; the receiver can reliably report a "collision" if r is outside the rate region. The maximum achievable communication rate region $\mathcal{R}_c$ is given in [25] by

$$\mathcal{R}_c = \{r | r \in \{r_1, \cdots, r_M\}, r < I_r(X; Y)\}, \tag{3.4}$$

where $I_r(\cdot, \cdot)$ is the mutual information function of the channel input and output symbols, computed using input distribution $P_{X|r}$. We assume that the operation region is contained in the maximum achievable rate region, i.e. $\mathcal{R} \subset \mathcal{R}_c$.

Equation (3.3) only gives the asymptotic limits on the error probabilities. In the rest of this section, we derive an achievable error probability bound under the assumption of *finite* codeword length N.

Define the system error probability with codeword length N, denoted by $P_{es}^{(N)}$, as

$$P_{es}^{(N)} = \max\left\{ \max_{(w,r), r\in\mathcal{R}} P_e^{(N)}(w,r), \max_{(w,r), r\notin\mathcal{R}} \bar{P}_c^{(N)}(w,r) \right\}. \tag{3.5}$$

The following theorem gives an achievable upper bound on $P_{es}^{(N)}$.

Theorem 3.1.1. *Consider single-user random access communication over discrete-time memoryless channel $P_{Y|X}$. Assume random coding with input distributions $P_{X|r}$, defined for all $r \in \{r_1, \cdots, r_M\}$. Let $\mathcal{R} \subseteq \{r_1, \cdots, r_M\}$ be an operation region. Given a codeword length N, there exists a decoder whose system error probability $P_{es}^{(N)}$ is upper bounded by*

$$P_{es}^{(N)} \leq \max\left\{ \begin{array}{l} \max_{r\in\mathcal{R}} \left[\begin{array}{c} \sum_{\tilde{r}\in\mathcal{R}} \exp\{-NE_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}})\} \\ + \max_{\tilde{r}\notin\mathcal{R}} \exp\{-NE_i(r, P_{X|r}, P_{X|\tilde{r}})\} \end{array} \right], \\ \sum_{r\in\mathcal{R}} \max_{\tilde{r}\notin\mathcal{R}} \exp\{-NE_i(r, P_{X|r}, P_{X|\tilde{r}})\} \end{array} \right\}, \tag{3.6}$$

where $E_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}})$ and $E_i(r, P_{X|r}, P_{X|\tilde{r}})$ are given by

$$\begin{aligned} E_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}}) &= \max_{0<\rho\leq 1} -\rho\tilde{r} + \max_{0<s\leq 1} \\ &\quad -\log\sum_Y \left[\sum_X P_{X|r}(X)P_{Y|X}(Y|X)^{1-s}\right]\left[\sum_X P_{X|\tilde{r}}(X)P_{Y|X}(Y|X)^{\frac{s}{\rho}}\right]^{\rho}, \\ E_i(r, P_{X|r}, P_{X|\tilde{r}}) &= \max_{0<\rho\leq 1} -\rho r + \max_{0<s\leq 1-\rho} \\ &\quad -\log\sum_Y \left[\sum_X P_{X|r}(X)P_{Y|X}(Y|X)^{\frac{s}{s+\rho}}\right]^{s+\rho}\left[\sum_X P_{X|\tilde{r}}(X)P_{Y|X}(Y|X)\right]^{1-s}. \end{aligned} \tag{3.7}$$

The proof of Theorem 3.1.1 is given in Section 3.6.1. In the proof, we assumed the following decoding algorithm at the receiver to achieve the error probability bound given in (3.6). Upon receiving the channel output symbols $\boldsymbol{y}$, the receiver outputs an estimated message and rate pair (w,r) with $r \in \mathcal{R}$ if the following condition is satisfied,

$$\begin{aligned} &\text{C1:} \ -\frac{1}{N}\log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(w,r)}\} < -\frac{1}{N}\log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})}\}, \\ &\qquad\qquad \text{for all } (\tilde{w},\tilde{r}) \neq (w,r), r, \tilde{r} \in \mathcal{R}, \\ &\text{C2:} \ -\frac{1}{N}\log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(w,r)}\} < \tau_r(\boldsymbol{y}), \end{aligned} \tag{3.8}$$

where $\tau_r(\cdot)$ is a pre-determined function of the channel output $\boldsymbol{y}$, associated with codewords of rate r. We term $\tau_r(\cdot)$ a typicality threshold function. If there is no codeword satisfying (3.8), the receiver reports a collision. In other words, the receiver decodes only if the log-likelihood of the maximum likelihood estimation exceeds certain threshold. Note that the random access codebook used to encode the message contains a large number of codewords, but the receiver only searches codewords corresponding to rates inside the operation region.

Define the corresponding exponent as the system error exponent $E_s = \lim_{N\to\infty} -\frac{1}{N}\log P_{es}^{(N)}$. Theorem 3.1.1 implies the following achievable bound on E_s.

Corollary 3.1.2. *The system error exponent of single-user random access communication given in Theorem 3.1.1 is lower-bounded by*

$$E_s = \lim_{N\to\infty} -\frac{1}{N}\log P_{es}^{(N)} \geq \min\left\{ \min_{r,\tilde{r}\in\mathcal{R}} E_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}}), \min_{r\in\mathcal{R},\tilde{r}\notin\mathcal{R}} E_i(r, P_{X|r}, P_{X|\tilde{r}}) \right\}, \tag{3.9}$$

where $E_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}})$ and $E_i(r, P_{X|r}, P_{X|\tilde{r}})$ are defined in (3.7).

Corollary 3.1.2 is implied by Theorem 3.1.1. An alternative proof is given in [52].

Note that if we define the decoding error exponent E_d and the collision miss detection exponent E_c as

$$\begin{aligned} E_d &= \min_{(w,r),r\in\mathcal{R}} \lim_{N\to\infty} -\frac{1}{N}\log P_e^{(N)}(w,r), \\ E_c &= \min_{(w,r),r\notin\mathcal{R}} \lim_{N\to\infty} -\frac{1}{N}\log \bar{P}_c^{(N)}(w,r), \end{aligned} \tag{3.10}$$

then the system error exponent equals the minimum of the two exponents, i.e., $E_s = \min\{E_d, E_c\}$. The lower bound of E_s given in (3.9) is obtained by optimizing the typicality threshold function $\tau_r(\cdot)$ as done in the proof of Theorem 3.1.1. It is easy to see that, for each $\boldsymbol{y}$, the decoding error exponent E_d increases in $\tau_r(\boldsymbol{y})$, while the collision miss detection exponent E_c decreases in $\tau_r(\boldsymbol{y})$. Therefore, $\tau_r(\cdot)$ can be used to adjust the tradeoff between E_d and E_c.

Also note that the first term on the right hand side of (3.9) corresponds to the maximum likelihood decoding criterion C1 in (3.8). This term becomes Gallager's random-coding exponent [4] if the input distributions associated to all rates are identical. The second term is due to the typical sequence decoding criterion C2 in (3.8). The two criteria, in conjunction, enabled collision detection at the receiver with a good decoding error performance.

We end this section by pointing out that the probability bound given in (3.6) can be further tightened, especially when the input distributions corresponding to $r \in \mathcal{R}$ are similar to each other. In the special case if the input distributions are identical for all rates, then the term $\sum_{\tilde{r} \in \mathcal{R}} \exp\{-NE_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}})\}$ in (3.6), which corresponds to the maximum likelihood decoding criterion C1 in (3.8), can be further improved to Gallager's bound given in [4][5]. However, in a general case, such improvement makes the error bound less structured comparing to (3.6), and it gives the same error exponent results. Therefore, we choose to skip the detailed discussion.

3.2 Random Multiple Access System

In this section, we consider K-user time-slotted random multiple access communication over a discrete-time memoryless channel. The channel is modeled by a conditional distribution $P_{Y|X_1,\cdots,X_K}$, where $X_k \in \mathcal{X}_k$ ($k \in \{1, \cdots, K\}$) is the channel input symbol of user k with $\mathcal{X}_k$ being the the finite input alphabet, and $Y \in \mathcal{Y}$ is the channel output symbol with $\mathcal{Y}$ being the finite output alphabet. Assume that the slot length equals N symbol durations, which is also the length of a packet. We again focus on coding within one time slot.

Suppose that at the beginning of a time slot, each user, say user k, chooses an arbitrary communication rate r_k, in nats per symbol, and encodes $\lfloor Nr_k \rfloor$ data nats, denoted by a message w_k, into a packet of N symbols. Assume $r_k \in \{r_{k_1}, \cdots, r_{k_M}\}$, where $\{r_{k_1}, \cdots, r_{k_M}\}$ is a predetermined set of rates, with cardinality M, known at the receiver. We assume the actual communication rates of the users are shared neither among each other, nor with the receiver. Whether the channel is known at the users (transmitters) is not important at this point. Because the global rate

[5] Specifically, we mean the bound given by (18) in [4] with $R = \frac{1}{N} \log \sum_{\tilde{r} \in \mathcal{R}} e^{N\tilde{r}}$.

information is not available, an individual user cannot know a priori whether or not its rate is supported by the channel in terms of reliable message recovery. Encoding is done using a random coding scheme described as follows. Let $\mathcal{L}_k = \{C_{k\theta_k} : \theta_k \in \Theta_k\}$ be a codebook library of user k, the codebooks of which are indexed by set Θ_k. Each codebook contains M classes of codewords. The i^{th} codeword class contains $\lfloor e^{Nr_{k_i}} \rfloor$ codewords, each with N symbols. Denote $C_{k\theta_k}(w_k, r_k)_j$ as the j^{th} symbol of the codeword corresponding to message w_k and communication rate r_k in codebook $C_{k\theta_k}$. User k first generates θ_k according to a distribution γ_k, such that random variables $X_{(w_k,r_k),j} : \theta_k \to C_{k\theta_k}(w_k, r_k)_j$ are independently distributed according to an input distribution $P_{X|r_k}$. User k then uses codebook $C_{k\theta_k}$ to map (w_k, r_k) into a codeword, denoted by $\boldsymbol{x}_{(w_k,r_k)}$, and sends it to the receiver.

Assume that the receiver knows the channel $P_{Y|X_1,\cdots,X_K}$ and the randomly generated codebooks of all users. Based on the channel and the codebook information, the receiver predetermines an "operation region" $\mathcal{R}$, which is a set of communication rate vectors under which the receiver *intends* to decode the messages. In each time slot, upon receiving the channel output symbol vector $\boldsymbol{y}$, the receiver outputs the estimated message and rate vector pair $(\hat{\boldsymbol{w}}, \hat{\boldsymbol{r}})$ (that contains the estimates for all users) only if $\hat{\boldsymbol{r}} \in \mathcal{R}$ and a predetermined decoding error probability requirement is satisfied. Otherwise the receiver outputs a collision.

To simplify the notations, we will use bold font vector variables to denote the corresponding variables of multiple users. For example, $\hat{\boldsymbol{w}}$ denotes the message estimates of all users, $\boldsymbol{r}$ denotes the communication rates of all users, $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$ denotes the input distributions conditioned on communication rates $\boldsymbol{r}$, etc. For a vector variable $\boldsymbol{r}$, we will use r_k to denote the element corresponding to user k. Let $\mathcal{S} \subset \{1, \cdots, K\}$ be an arbitrary subset of user indices. We will use $\boldsymbol{r}_{\mathcal{S}}$ to denote the communication rates of users in $\mathcal{S}$, and will use $\boldsymbol{w}_{\bar{\mathcal{S}}}$ to denote the messages of users not in $\mathcal{S}$, etc.

Similar to the single-user system, conditioned on $(\boldsymbol{w}, \boldsymbol{r})$ is transmitted, we define the decoding error probability for $(\boldsymbol{w}, \boldsymbol{r})$ with $\boldsymbol{r} \in \mathcal{R}$ and codeword length N as

$$P_e^{(N)}(\boldsymbol{w}, \boldsymbol{r}) = Pr\{(\hat{\boldsymbol{w}}, \hat{\boldsymbol{r}}) \neq (\boldsymbol{w}, \boldsymbol{r}) | (\boldsymbol{w}, \boldsymbol{r})\}, \forall (\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}. \tag{3.11}$$

We define the collision miss detection probabilities for $(\boldsymbol{w}, \boldsymbol{r})$ with $\boldsymbol{r} \notin \mathcal{R}$ and codeword length N as

$$\bar{P}_c^{(N)}(\boldsymbol{w}, \boldsymbol{r}) = 1 - Pr\{\text{``collision''}|(\boldsymbol{w}, \boldsymbol{r})\}, \forall(\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \notin \mathcal{R}. \tag{3.12}$$

A rate region $\mathcal{R}$ is said to be *achievable* if there exists a series of random coding schemes $(\mathcal{L}^{(N)}, \gamma^{(N)})$ with the decoding error probability and collision miss detection probability defined in (3.11) and (3.12), such that

$$\begin{aligned} &\lim_{N\to\infty} P_e^{(N)}(\boldsymbol{w}, \boldsymbol{r}) = 0, \qquad \forall(\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}, \\ &\lim_{N\to\infty} \bar{P}_c^{(N)}(\boldsymbol{w}, \boldsymbol{r}) = 0, \qquad \forall(\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \notin \mathcal{R}. \end{aligned} \tag{3.13}$$

In other words, asymptotically as N goes to infinity, the receiver can reliably decode the messages for all rate vectors inside $\mathcal{R}$ and can reliably report a collision for all rate vectors outside $\mathcal{R}$. It has been proved in [25] that the maximum achievable communication rate region for such multiple random access system $\mathcal{R}_c$ is given by

$$\mathcal{R}_c = \left\{ \boldsymbol{r} \;\middle|\; \begin{array}{l} r_k \in \{r_{k1}, \cdots, r_{kM}\}, k \in \{1, \cdots, K\} \\ \forall \mathcal{S} \subset \{1, \cdots, K\}, \sum_{k \notin \mathcal{S}} r_k < I_{\boldsymbol{r}}(\boldsymbol{X}_{\bar{\mathcal{S}}}; Y|\boldsymbol{X}_{\mathcal{S}}) \end{array} \right\}, \tag{3.14}$$

where $I_{\boldsymbol{r}}(\boldsymbol{X}_{\bar{\mathcal{S}}}; Y|\boldsymbol{X}_{\mathcal{S}})$ is the conditional mutual information computed using input distribution $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$. We still assume the operation region satisfies $\mathcal{R} \subset \mathcal{R}_c$.

Define the system error probability $P_{es}^{(N)}$, with codeword length N, as

$$P_{es}^{(N)} = \max\left\{ \max_{(\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}} P_e^{(N)}(\boldsymbol{w}, \boldsymbol{r}), \max_{(\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \notin \mathcal{R}} \bar{P}_c^{(N)}(\boldsymbol{w}, \boldsymbol{r}) \right\}. \tag{3.15}$$

The following theorem gives an upper bound on $P_{es}^{(N)}$.

Theorem 3.2.1. *For K-user random multiple access communication over a discrete time memoryless channel $P_{Y|\boldsymbol{X}}$. Assume finite codeword length N, and random coding with input distribution $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$ for all $\boldsymbol{r}$ with $r_k \in \{r_{k_1}, \cdots, r_{k_M}\}$, $1 \le k \le K$. Let $\mathcal{R}$ be the operation region. There exists* *a decoding algorithm, whose system error probability $P_{es}^{(N)}$ is upper bounded by*

$$P_{es}^{(N)} \leq \max \Bigg\{ \max_{\boldsymbol{r} \in \mathcal{R}} \sum_{\mathcal{S} \subset \{1, \cdots, K\}} \Bigg[\sum_{\tilde{\boldsymbol{r}} \in \mathcal{R}, \tilde{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp\{-N E_m(\mathcal{S}, \tilde{\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}})\}$$

$$+ \max_{\boldsymbol{r}' \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp\{-N E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})\} \Bigg],$$

$$\max_{\tilde{\boldsymbol{r}} \notin \mathcal{R}} \sum_{\mathcal{S} \subset \{1, \cdots, K\}} \sum_{\boldsymbol{r} \in \mathcal{R}, \boldsymbol{r}_{\mathcal{S}} = \tilde{\boldsymbol{r}}_{\mathcal{S}}} \max_{\boldsymbol{r}' \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \tilde{\boldsymbol{r}}_{\mathcal{S}}} \exp\{-N E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})\} \Bigg\}, \tag{3.16}$$

where $E_m(\mathcal{S}, \tilde{\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}})$ and $E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})$ are given by

$$E_m(\mathcal{S}, \tilde{\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}}) = \max_{0 < \rho \leq 1} -\rho \sum_{k \notin \mathcal{S}} \tilde{r}_k + \max_{0 < s \leq 1} -\log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k \in \mathcal{S}} P_{X|r_k}(X_k)$$

$$\times \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{1-s} \right) \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|\tilde{r}_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{\rho}} \right)^{\rho},$$

$$E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'}) = \max_{0 < \rho \leq 1} -\rho \sum_{k \notin \mathcal{S}} r_k + \max_{0 < s \leq 1-\rho} -\log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k \in \mathcal{S}} P_{X|r_k}(X_k)$$

$$\times \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{s+\rho}} \right)^{s+\rho} \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r'_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X}) \right)^{1-s}. \tag{3.17}$$

The proof of Theorem 3.2.1 is given in Section 3.6.2. In the proof, we assumed the following decoding algorithm at the receiver to achieve the error probability bound given in (3.16). Upon receiving the channel output symbols $\boldsymbol{y}$, the receiver outputs an estimated message vector and rate vector pair $(\boldsymbol{w}, \boldsymbol{r})$ with $\boldsymbol{r} \in \mathcal{R}$ if both the following two conditions are satisfied.

$$\text{C1: } -\frac{1}{N} \log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}\} < -\frac{1}{N} \log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}\},$$

$$\text{for all } (\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}}) \neq (\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r}, \tilde{\boldsymbol{r}} \in \mathcal{R},$$

$$\text{C2: } -\frac{1}{N} \log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}\} < \tau_{\boldsymbol{r}}(\boldsymbol{y}), \tag{3.18}$$

where $\tau_{\boldsymbol{r}}(\cdot)$ is a pre-determined typicality threshold function of the channel output $\boldsymbol{y}$, associated with codewords of rate $\boldsymbol{r}$. If there is no codeword satisfying (3.18), the receiver reports a collision.

Define the corresponding exponent as the system error exponent $E_s = \lim_{N \to \infty} -\frac{1}{N} \log P_{es}^{(N)}$. Theorem 3.2.1 implies the following achievable bound on E_s.

Corollary 3.2.2. *The system error exponent of the random multiple access communication system given in Theorem 3.2.1is lower-bounded by*

$$E_s \geq \min\left\{ \min_{\mathcal{S}\subset\{1,\cdots,K\}} \min_{\boldsymbol{r},\tilde{\boldsymbol{r}}\in\mathcal{R},\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}} E_m(\mathcal{S},\tilde{\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}}), \right.$$
$$\left. \min_{\mathcal{S}\subset\{1,\cdots,K\}} \min_{\boldsymbol{r}\in\mathcal{R},\tilde{\boldsymbol{r}}\notin\mathcal{R},\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}} E_i(\mathcal{S},\boldsymbol{r},\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}}) \right\}, \tag{3.19}$$

where $E_m(\mathcal{S},\tilde{\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}})$ and $E_i(\mathcal{S},\boldsymbol{r},\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}})$ are defined in (3.17).

Corollary 3.2.2 is implied by Theorem 3.2.1.

Similarly to the single-user system, if we define the decoding error exponent E_d and the collision miss detection exponent E_c as

$$E_d = \min_{(\boldsymbol{w},\boldsymbol{r}),\boldsymbol{r}\in\mathcal{R},} \lim_{N\to\infty} -\frac{1}{N}\log P_e^{(N)}(\boldsymbol{w},\boldsymbol{r}),$$
$$E_c = \min_{(\boldsymbol{w},\boldsymbol{r}),\boldsymbol{r}\notin\mathcal{R}} \lim_{N\to\infty} -\frac{1}{N}\log \bar{P}_c^{(N)}(\boldsymbol{w},\boldsymbol{r}), \tag{3.20}$$

then the system error exponent equals the minimum of the two exponents, i.e., $E_s = \min\{E_d, E_c\}$. Again, instead of optimizing the typicality function $\tau_{\boldsymbol{r}}(\cdot)$ to lower bound E_s, $\tau_{\boldsymbol{r}}(\cdot)$ can be used to adjust the tradeoff between E_d and E_c.

3.3 Error Performance under Generalized Random Coding

In the previous sections, we used the practical definition of communication rate, i.e., communication rate equals the normalized data nats per symbol encoded in a packet. Codewords of each user are partitioned into M classes each corresponding to a rate option. This is equivalent to indexing the codewords using a message and rate pair (w, r). We assumed codeword symbols within each class, i.e., corresponding to the same r, should be randomly generated according to the same input distribution. In this section, we extend the results to the generalized random coding scheme [25] where symbols of different codewords, as opposed to different codeword classes, can be generated according to different input distributions.

We will index the codewords in a codebook using a macro message W, which is essentially another expression of the (w, r) pair used in previous sections. In other words, W contains both information about the message w and the rate r in practical senses. The generalized random coding scheme is defined originally in [25] as follows.

Definition 3.3.1. *(***generalized random coding** *[25]) Let $\mathcal{L} = \{C_\theta : \theta \in \Theta\}$ be a library of codebooks. Each codebook in the library contains $e^{NR_{\max}}$ codewords of length N, where $R_{\max}$ is an arbitrary large finite constant. Let the codebooks be indexed by a set Θ. Let the actual codebook chosen by the transmitter be C_θ where the index θ is a random variable following distribution γ. Let $W \in \{1, \cdots, e^{NR_{\max}}\}$ be a macro message used to index the codewords in each codebook. Denote $C_\theta(W)_j$ as the j^{th} symbol of the codeword corresponding to macro message W in codebook C_θ. We define $(\mathcal{L}, \gamma)$ as a generalized random coding scheme following distribution $P_{X|W}$, if the random variables $X_{W,j} : \theta \to C_\theta(W)_j$, $\forall j, W$, are independently distributed according to input distribution $P_{X|W}$.*

Note that a generalized random coding scheme allows codeword symbols corresponding to different messages to be generated according to different input distributions. Because codewords are indexed using macro message W, communication rate r becomes a function of W. Consequently, the practical communication rate r used in previous sections only represents a specific choice of the rate function. In order to distinguish codewords from each other in rate and error performance characterization, in this section, we will switch to the following *standard communication rate* definition, originally introduced in [25].

Definition 3.3.2. *(***standard communication rate** *[25]) Assume codebook C has $e^{NR_{\max}}$ codewords of length N, where $R_{\max}$ is an arbitrary large finite constant. Let the corresponding messages or codewords be indexed by $W \in \{1, \cdots, e^{NR_{\max}}\}$. For each message W, we define its standard communication rate, in nats per symbol, as $r(W) = \frac{1}{N} \log W$.*

Since the standard rate function $r(W) = \frac{1}{N}\log W$ is invertible, system performance characterized in any other rate function can be derived from that of the standard rate function [25][6].

The following definition specifies a sequence of generalized random coding schemes following an asymptotic input distribution.

Definition 3.3.3. *(***asymptotic input distribution** *[25]) Let $\{(\mathcal{L}^{(N)}, \gamma^{(N)})\}$ be a sequence of random coding schemes, where $(\mathcal{L}^{(N)}, \gamma^{(N)})$ is a generalized random coding scheme with codeword length N and input distribution $P^{(N)}_{X|W^{(N)}}$. Assume each codebook in library $\mathcal{L}^{(N)}$ has $e^{NR_{\max}}$ codewords. Let $P_{X|r}$ be an input distribution defined as a function of the standard rate r, for all $r \in [0, R_{\max}]$. We say $\{(\mathcal{L}^{(N)}, \gamma^{(N)})\}$ follows an asymptotic input distribution $P_{X|r}$, if for all $\{W^{(N)}\}$ sequences with well defined rate limit $\lim_{N\to\infty} r(W^{(N)})$, we have*

$$\lim_{N\to\infty} P^{(N)}_{X|W^{(N)}} = \lim_{N\to\infty} P_{X|r(W^{(N)})}. \tag{3.21}$$

Note that since we do not assume $P_{X|r}$ is continuous in r, we may not have $\lim_{N\to\infty} P_{X|r(W^{(N)})} = P_{X|\lim_{N\to\infty} r(W^{(N)})}$.

Let us still use bold font vector variables to denote the corresponding variables of multiple users. Theorem 3.3.4 gives the achievable error exponent of a random multiple access system using generalized random coding.

Theorem 3.3.4. *Consider K-user random multiple access communication over a discrete-time memoryless channel $P_{Y|\boldsymbol{X}}$ using a sequence of generalized random coding schemes, denoted by $\{(\boldsymbol{\mathcal{L}}^{(N)}, \boldsymbol{\gamma}^{(N)})\}$. Assume that $\{(\boldsymbol{\mathcal{L}}^{(N)}, \boldsymbol{\gamma}^{(N)})\}$ follows asymptotic distribution $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$. For any user k, assume $P_{X_k|r_k}$ is only discontinuous in r_k at a finite number of points. Let the operation region $\mathcal{R}$ be strictly contained in an achievable rate region, specified in [25]. Equation (3.19) gives an achievable lower bound on the system error exponent E_s, with rates in the equation being the standard communication rates.*

[6] Note that the standard rate is defined here using the natural logarithm, while it was defined using the base-2 logarithm in [25].

The proof of Theorem 3.3.4 is given in Section 3.6.3. In the proof, an achievable error probability bound in the case of a finite codeword length is also given in Lemma 3.6.1.

3.4 Random Access Communication over Compound Channels

In this section, we relax the assumption that the channel state information is known at the receiver. We assume that the random multiple access communication takes places over a compound channel, which consists of a family of discrete-time memoryless channels, characterized by a finite set of conditional probabilities $\left\{P^{(1)}_{Y|\boldsymbol{X}}, \cdots, P^{(H)}_{Y|\boldsymbol{X}}\right\}$ with cardinality H. For each time slot, a channel realization is randomly generated from this set and remains static within the slot duration. We assume that all users and the receiver know the compound channel set, but not the actual channel realization. For the time being, we will assume that $H < \infty$. The case when the compound channel set contains an infinite number of channels will be discussed at the end of this section.

Similarly as in Section 3.1 and Section 3.2, we still focus on channel coding within one time slot. The same random coding scheme is used to encode the transmitted messages of all users. We assume that the receiver is shared with the random codebook generation algorithms and hence knows the randomly generated codebooks of all users. Before packet transmission, the receiver pre-determines an "operation region" $\mathcal{R} = \{(\boldsymbol{r}, P_{Y|\boldsymbol{X}})\}$, which is a set of rate vector and channel realization pair, where each entry of $\boldsymbol{r}$ is chosen from the corresponding rate set, i.e., $r_k \in \{r_{k1}, \cdots, r_{kM}\}$ ($k \in \{1, \cdots, K\}$), and $P_{Y|\boldsymbol{X}} \in \left\{P^{(1)}_{Y|\boldsymbol{X}}, \cdots, P^{(H)}_{Y|\boldsymbol{X}}\right\}$. Let $(\boldsymbol{r}, P_{Y|\boldsymbol{X}})$ be the actual realization of the transmitted rate vector and channel pair. We assume that the receiver *intends* to decode all messages if $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}$, and *intends* to report a collision if $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \notin \mathcal{R}$. Note that the actual rate and channel realization $(\boldsymbol{r}, P_{Y|\boldsymbol{X}})$ is unknown at the receiver. Therefore the receiver needs to make decisions whether to decode messages or to report a collision only based on the received channel symbols. More specifically, in each time slot, upon receiving the channel output symbols $\boldsymbol{y}$, the receiver estimates the rate and channel pair, denoted by $(\hat{\boldsymbol{r}}, \hat{P}_{Y|\boldsymbol{X}})$, for all users. The receiver outputs the corresponding estimated message and rate vector pair $(\hat{\boldsymbol{w}}, \hat{\boldsymbol{r}})$ if $(\hat{\boldsymbol{r}}, \hat{P}_{Y|\boldsymbol{X}}) \in \mathcal{R}$ and a pre-determined decoding error probability requirement is satisfied. Otherwise, the receiver reports a collision. Also note that, whether the receiver should recover the messages or report a collision not only depends on the rates, but also depends on the channel realization. In other words, for the same transmission rate vector, the receiver may be designed to take different actions for different channel realizations. This is opposed to the conventional compound channel communication scenario where, if a rate is supported by the system, the receiver should always decode the messages irrespective of the channel realization.

Given the operation region $\mathcal{R}$, and conditioned on that $(\boldsymbol{w}, \boldsymbol{r})$ is transmitted over channel $P_{Y|\boldsymbol{X}}$, we define the following three error probabilities. The decoding error probability, for $(\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}})$ with $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}$, is defined as

$$P_{e(\boldsymbol{w},\boldsymbol{r},P_{Y|\boldsymbol{X}})} = Pr\left\{(\hat{\boldsymbol{w}}, \hat{\boldsymbol{r}}) \neq (\boldsymbol{w}, \boldsymbol{r}) | (\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}})\right\}, \quad \forall (\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}}), (\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}. \tag{3.22}$$

The collision miss detection probability, for $(\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}})$ with $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \notin \mathcal{R}$, is defined as

$$\begin{aligned}\bar{P}_{c(\boldsymbol{w},\boldsymbol{r},P_{Y|\boldsymbol{X}})} = 1 - Pr\left\{\text{“collision”} | (\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}})\right\} \\ - Pr\left\{(\hat{\boldsymbol{w}}, \hat{\boldsymbol{r}}) = (\boldsymbol{w}, \boldsymbol{r}) | (\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}})\right\}, \\ \forall (\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}}), (\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \notin \mathcal{R}.\end{aligned} \tag{3.23}$$

Note that in (3.23), when $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \notin \mathcal{R}$, we have excluded the correct message and rate pair estimation from the collision miss detection event.

Let $\mathcal{S} \subset \{1, \cdots, K\}$ be an arbitrary user subset. Assume that $\sum_{k \notin \mathcal{S}} r_k \leq I_{(\boldsymbol{r}, P_{Y|\boldsymbol{X}})}(\boldsymbol{X}_{\bar{\mathcal{S}}}; \boldsymbol{Y} | \boldsymbol{X}_{\mathcal{S}})$ for all $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}$, where $\boldsymbol{X}_{\mathcal{S}}$ denotes the channel input symbols of users in set $\mathcal{S}$, and $\boldsymbol{X}_{\bar{\mathcal{S}}}$ denotes the channel input symbols of users not in set $\mathcal{S}$. $I_{(\boldsymbol{r}, P_{Y|\boldsymbol{X}})}$ is the mutual information function computed using input distribution corresponding to rate vector $\boldsymbol{r}$ (i.e., $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$) and channel $P_{Y|\boldsymbol{X}}$. We define the system error probability P_{es} as

$$\begin{aligned}P_{es} = \max\Bigg\{ \max_{(\boldsymbol{w},\boldsymbol{r},P_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}) \in \mathcal{R}} P_{e(\boldsymbol{w},\boldsymbol{r},P_{Y|\boldsymbol{X}})}, \\ \max_{(\boldsymbol{w},\boldsymbol{r},P_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}) \notin \mathcal{R}} \bar{P}_{c(\boldsymbol{w},\boldsymbol{r},P_{Y|\boldsymbol{X}})} \Bigg\}.\end{aligned} \tag{3.24}$$

The following theorem gives an upper bound on the achievable system error probability P_{es}.

Theorem 3.4.1. *Consider K-user random multiple access communication over compound discrete-time memoryless channel $\left\{P_{Y|\boldsymbol{X}}^{(1)}, \cdots, P_{Y|\boldsymbol{X}}^{(H)}\right\}$, where $H < \infty$ is a positive integer. Let $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$ be the input distribution for all users and all rates. Let $\mathcal{R}$ be the operation region. Assume finite codeword length N. There exists a decoding algorithm, whose system error probability $P_{es}^{(N)}$ is upper bounded by,*

$$
\begin{aligned}
P_{es}^{(N)} \leq \max\Bigg\{ & \max_{(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}} \sum_{\mathcal{S} \subset \{1, \cdots, K\}} \\
& \Bigg[\sum_{(\tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}) \in \mathcal{R}, \tilde{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp\{-N E_m(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, P_{Y|\boldsymbol{X}}, \tilde{P}_{Y|\boldsymbol{X}})\} \\
& \quad + \max_{(\boldsymbol{r}', P'_{Y|\boldsymbol{X}}) \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp\{-N E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{r}', P_{Y|\boldsymbol{X}}, P'_{Y|\boldsymbol{X}})\} \Bigg], \\
& \max_{(\tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}) \notin \mathcal{R}} \sum_{\mathcal{S} \subset \{1, \cdots, K\}} \sum_{(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}, \boldsymbol{r}_{\mathcal{S}} = \tilde{\boldsymbol{r}}_{\mathcal{S}}} \\
& \quad \max_{(\boldsymbol{r}', P'_{Y|\boldsymbol{X}}) \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \tilde{\boldsymbol{r}}_{\mathcal{S}}} \exp\{-N E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{r}', P_{Y|\boldsymbol{X}}, P'_{Y|\boldsymbol{X}})\} \Bigg\},
\end{aligned}
\tag{3.25}
$$

where $E_m(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, P_{Y|\boldsymbol{X}}, \tilde{P}_{Y|\boldsymbol{X}})$ and $E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{r}', P_{Y|\boldsymbol{X}}, P'_{Y|\boldsymbol{X}})$ are given by

$$
\begin{aligned}
E_m(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, P_{Y|\boldsymbol{X}}, \tilde{P}_{Y|\boldsymbol{X}}) = & \max_{0 < \rho \leq 1} -\rho \sum_{k \notin \mathcal{S}} \tilde{r}_k + \max_{0 < s \leq 1} - \log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k \in \mathcal{S}} P_{X|r_k}(x_k) \\
\times \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{1-s} \right) & \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|\tilde{r}_k}(X_k) \tilde{P}_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{\rho}} \right)^{\rho}, \\
E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{r}', P_{Y|\boldsymbol{X}}, P'_{Y|\boldsymbol{X}}) = & \max_{0 < \rho \leq 1} -\rho \sum_{k \notin \mathcal{S}} r_k + \max_{0 < s \leq 1-\rho} - \log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k \in \mathcal{S}} P_{X|r_k}(X_k) \\
\times \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{s+\rho}} \right)^{s+\rho} & \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r'_k}(X_k) P'_{Y|\boldsymbol{X}}(Y|\boldsymbol{X}) \right)^{1-s}.
\end{aligned}
\tag{3.26}
$$

The proof of Theorem 3.4.1 can is given in Section 3.6.5.

When the compound channel is randomly generated at the beginning but remains static afterwards, one can take codeword length to infinity to obtain the system error exponent as $E_s = \lim_{N \to \infty} -\frac{1}{N} \log P_{es}$. The following lower bound on the achievable system error exponent E_s can be easily derived from Theorem 3.4.1.

Corollary 3.4.2. *The system error exponent of a K-user multiple random access system over compound discrete-time memoryless channels given in Theorem 3.4.1 is lower-bounded by*

$$E_s \geq \min \left\{ \min_{(\boldsymbol{r}, P_{Y|\boldsymbol{X}}), (\tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}) \in \mathcal{R},} E_m(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, P_{Y|\boldsymbol{X}}, \tilde{P}_{Y|\boldsymbol{X}}), \right.$$
$$\left. \min_{(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}, (\tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}) \notin \mathcal{R},} E_i(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, P_{Y|\boldsymbol{X}}, \tilde{P}_{Y|\boldsymbol{X}}) \right\}, \tag{3.27}$$

where $E_m(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, P_{Y|\boldsymbol{X}}, \tilde{P}_{Y|\boldsymbol{X}})$ and $E_i(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, P_{Y|\boldsymbol{X}}, \tilde{P}_{Y|\boldsymbol{X}})$ are given in (3.26).

Compared with the error exponent derived in Corollary 3.2.2, we can see that, even though the channel stays static forever, the system still needs to pay a penalty in error exponent performance for not knowing the channel at the receiver[7].

In both Theorem 3.4.1 and Corollary 3.4.2, we have assumed that there are only a finite number of channels in the compound set. Next, we will extend the result to the case when the cardinality of the compound channel set can be infinity.

We first assume that the the channels in the compound set can be partitioned into H classes, denoted by $\left\{\mathcal{F}^{(1)}, \cdots, \mathcal{F}^{(H)}\right\}$, where $H < \infty$ is a positive integer. For example, if the compound channel set contains fading channels with continuous channel gains, one could quantize the channel gains and define the set of channels with the same quantization outcome as one channel class. We next assume that the receiver should choose an operation region $\mathcal{R}$ to satisfy the following constraint for any rate vector $\boldsymbol{r}$ and channel class $\mathcal{F} \in \left\{\mathcal{F}^{(1)}, \cdots, \mathcal{F}^{(H)}\right\}$.

$$\text{C1: For any } (\boldsymbol{r}, \mathcal{F}), \text{ either } (\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R} \ \forall P_{Y|\boldsymbol{X}} \in \mathcal{F},$$
$$\text{or } (\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \notin \mathcal{R} \ \forall P_{Y|\boldsymbol{X}} \in \mathcal{F}. \tag{3.28}$$

We say $(\boldsymbol{r}, \mathcal{F}) \in \mathcal{R}$ if $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}$ for all $P_{Y|\boldsymbol{X}} \in \mathcal{F}$, and we say $(\boldsymbol{r}, \mathcal{F}) \notin \mathcal{R}$ otherwise.

For each channel class $\mathcal{F}$ and for each channel output symbol Y and input symbol vector $\boldsymbol{X}$, we define the following upper and lower bounds on the conditional probability values yielded by

[7] We assume that such a conclusion should be well known for the conventional compound channel communication. However, we are not able to find a reference that made such a clear statement.

channels in $\mathcal{F}$, denoted by $P_{\max}^{\mathcal{F}}(Y|\boldsymbol{X})$ and $P_{\min}^{\mathcal{F}}(Y|\boldsymbol{X})$,

$$\begin{aligned}
P_{\max}^{\mathcal{F}}(Y|\boldsymbol{X}) &= \max_{P_{Y|\boldsymbol{X}}\in\mathcal{F}} P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X}),\\
P_{\min}^{\mathcal{F}}(Y|\boldsymbol{X}) &= \min_{P_{Y|\boldsymbol{X}}\in\mathcal{F}} P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X}).
\end{aligned} \tag{3.29}$$

The following theorem gives an upper bound on the achievable system error probability.

Theorem 3.4.3. *Consider a K-user multiple random access communication system over a compound discrete-time memoryless channel. Assume that the compound set is partitioned into H classes, denoted by $\left\{\mathcal{F}^{(1)}, \cdots, \mathcal{F}^{(H)}\right\}$, where H is a finite positive integer. Assume that the operation region $\mathcal{R}$ satisfies constraint C1 given in (3.28). The system error probability P_{es} is upper bounded as follows.*

$$\begin{aligned}
P_{es} \le \max\Bigg\{ &\max_{(\boldsymbol{r},\mathcal{F})\in\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \Bigg[\max_{(\boldsymbol{r}',\mathcal{F}')\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \exp\left\{-NE_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',\mathcal{F},\mathcal{F}')\right\}\\
&\qquad + \sum_{(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}})\in\mathcal{R},\tilde{\boldsymbol{r}}_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \exp\left\{-NE_m(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{F},\tilde{\mathcal{F}})\right\}\Bigg],\\
&\max_{(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}})\notin\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \Bigg[\sum_{(\boldsymbol{r},\mathcal{F})\in\mathcal{R},\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}} \max_{(\boldsymbol{r}',\mathcal{F}')\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \exp\left\{-NE_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',\mathcal{F},\mathcal{F}')\right\}\Bigg]\Bigg\}.
\end{aligned} \tag{3.30}$$

where $E_m(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{F},\tilde{\mathcal{F}})$ and $E_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',\mathcal{F},\mathcal{F}')$ are given by

$$\begin{aligned}
E_m(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{F},\tilde{\mathcal{F}}) &= \max_{0<\rho\le 1} -\rho\sum_{k\notin\mathcal{S}}\tilde{r}_k + \max_{0<s\le 1} -\log\sum_{Y}\sum_{\boldsymbol{X}_{\mathcal{S}}}\prod_{k\in\mathcal{S}} P_{X|r_k}(X_k)\\
&\times\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|r_k}(X_k)P_{\max}^{\mathcal{F}}(Y|\boldsymbol{X})P_{\min}^{\mathcal{F}}(Y|\boldsymbol{X})^{-s}\right)\\
&\times\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|\tilde{r}_k}(X_k)P_{\max}^{\tilde{\mathcal{F}}}(Y|\boldsymbol{X})^{\frac{s}{\rho}}\right)^{\rho}.\\
E_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',\mathcal{F},\mathcal{F}') &= \max_{0<\rho\le 1} -\rho\sum_{k\notin\mathcal{S}} r_k + \max_{0<s\le 1-\rho} -\log\sum_{Y}\sum_{\boldsymbol{X}_{\mathcal{S}}}\prod_{k\in\mathcal{S}} P_{X|r_k}(X_k)\\
&\times\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|r_k}(X_k)P_{\max}^{\mathcal{F}}(Y|\boldsymbol{X})P_{\min}^{\mathcal{F}}(Y|\boldsymbol{X})^{\frac{-\rho}{s+\rho}}\right)^{s+\rho}\\
&\times\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|r'_k}(X_k)P_{\max}^{\mathcal{F}'}(Y|\boldsymbol{X})\right)^{1-s}.
\end{aligned} \tag{3.31}$$

The proof of Theorem 3.4.3 is given in Section 3.6.6. As shown in the proof that, in order to make decoding and collision report decisions, the receiver only needs to search over the finite number of channel classes using statistics $P^{\mathcal{F}}_{\max}$ and $P^{\mathcal{F}}_{\min}$ defined in (3.29), as opposed to searching among all possible channels.

3.5 Individual User Decoding

In Section 3.2, we have assumed that the receiver either decodes messages or reports collisions for *all* users in the system. In practical applications, even though many users compete for the wireless channel, it is common that the receiver may not be interested in recovering messages for all of them. In this section, we show that the results obtained in Section 3.2 can help to derive error probability bounds in a random multiple access system where the receiver is only interested in recovering the messages from a user subset. However, to simplify the notations, we will only consider a special case when the communication channel is known at the receiver, and when the receiver is only interested in decoding for a single user. Generalizing the results to decoding for multiple users over a compound channel is straightforward.

Let the discrete-time memoryless channel be characterized by $P_{Y|\boldsymbol{X}}$, which is known at the receiver. In each time slot, each user chooses a communication rate and encodes its message using the random coding scheme described in Section 3.2. The rate information is shared neither among the users nor with the receiver. We assume that the receiver is only interested in recovering the message for user $k \in \{1, \cdots, K\}$. We assume that the receiver chooses an operation region $\mathcal{R}$, such that if the transmitted rate vector $\boldsymbol{r}$ satisfies $\boldsymbol{r} \in \mathcal{R}$, the receiver intends to decode for user k, and if $\boldsymbol{r} \notin \mathcal{R}$, the receiver intends to report a collision for user k. It is important to note that, first, whether the receiver will be able to decode the message of user k, not only depends on the rate of user k, but also depends on the rate of other users. Therefore, the operation rate region $\mathcal{R}$ should still be defined as a set of rate vector $\boldsymbol{r}$, as opposed to the rate of user k. Second, even though the receiver only cares about the message of user k, the receiver still has the option of decoding the messages for some other users if this helps to improve the communication performance of user k.

This implies that, based upon the received symbols, the receiver will essentially need to make a decision on which subset of the messages should be decoded.

Due to the above understandings, we first define an elementary decoder, called the "$(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder". Given a user subset $\mathcal{D} \subseteq \{1, \cdots, K\}$ and an operation rate region $\mathcal{R}_{\mathcal{D}}$, the "$(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder" intends to recover messages for users in $\mathcal{D}$ while regarding signals from users not in $\mathcal{D}$ as interference, if the communication rate vector is within the operation region $\mathcal{R}_{\mathcal{D}}$. We define the following error probabilities for a $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder. Conditioned on users in $\mathcal{D}$ transmitting $(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}})$ and users not in $\mathcal{D}$ choosing rate $\boldsymbol{r}_{\bar{\mathcal{D}}}$, let us denote the estimated messages and rates by $(\hat{\boldsymbol{w}}_{\mathcal{D}}, \hat{\boldsymbol{r}}_{\mathcal{D}})$ and $\hat{\boldsymbol{r}}_{\bar{\mathcal{D}}}$ if the decoder does not report a collision. We define the decoding error probability of the $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder for $(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}})$ with $\boldsymbol{r} \in \mathcal{R}_{\mathcal{D}}$ as

$$\begin{aligned} P_e(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}) = Pr\{(\hat{\boldsymbol{w}}_{\mathcal{D}}, \hat{\boldsymbol{r}}_{\mathcal{D}}) \neq (\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}) | (\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}})\}, \\ \forall (\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}), \boldsymbol{r} \in \mathcal{R}_{\mathcal{D}}. \end{aligned} \tag{3.32}$$

We define the collision miss detection probability for $(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}})$ with $\boldsymbol{r} \notin \mathcal{R}_{\mathcal{D}}$ as

$$\begin{aligned} \bar{P}_c(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}) = 1 - Pr\{\text{“collision”} | (\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}})\} \\ - Pr\{(\hat{\boldsymbol{w}}_{\mathcal{D}}, \hat{\boldsymbol{r}}_{\mathcal{D}}) = (\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}) | (\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}})\}, \\ \forall (\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}), \boldsymbol{r} \notin \mathcal{R}_{\mathcal{D}}. \end{aligned} \tag{3.33}$$

System error probability of the $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder is defined by

$$\begin{aligned} P_{es}(\mathcal{D}, \mathcal{R}_{\mathcal{D}}) = \max \Bigg\{ \max_{(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}), \boldsymbol{r} \in \mathcal{R}_{\mathcal{D}}} P_e(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}), \\ \max_{(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}), \boldsymbol{r} \notin \mathcal{R}_{\mathcal{D}}} \bar{P}_c(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{r}_{\mathcal{D}}, \boldsymbol{r}_{\bar{\mathcal{D}}}) \Bigg\}. \end{aligned} \tag{3.34}$$

Given a finite codeword length N, the following lemma gives an upper bound on the achievable system error probability of a $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder.

Lemma 3.5.1. *The following system error probability bound is achievable for a K-user random multiple access communication system over a discrete-time memoryless channel $P_{Y|\boldsymbol{X}}$ with an*

$(\mathcal{D},\mathcal{R}_{\mathcal{D}})$-decoder,

$$
\begin{aligned}
P_{es}(\mathcal{D},\mathcal{R}_{\mathcal{D}}) \leq \max \Bigg\{ & \max_{\boldsymbol{r}\in\mathcal{R}_{\mathcal{D}}} \sum_{\mathcal{S}\subset\mathcal{D}} \Bigg[\sum_{\substack{\tilde{\boldsymbol{r}}\,\in\,\mathcal{R}_{\mathcal{D}},\\ \tilde{\boldsymbol{r}}_{\mathcal{S}}\,=\,\boldsymbol{r}_{\mathcal{S}}}} \exp\{-NE_{m\mathcal{D}}(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}})\} \\
& \qquad\qquad + \max_{\substack{\boldsymbol{r}'\,\notin\,\mathcal{R}_{\mathcal{D}},\\ \boldsymbol{r}'_{\mathcal{S}}\,=\,\boldsymbol{r}_{\mathcal{S}}}} \exp\{-NE_{i\mathcal{D}}(\mathcal{S},\boldsymbol{r},\boldsymbol{r}')\} \Bigg], \\
& \max_{\tilde{\boldsymbol{r}}\notin\mathcal{R}_{\mathcal{D}}} \sum_{\mathcal{S}\subset\mathcal{D}} \sum_{\substack{\boldsymbol{r}\,\in\,\mathcal{R}_{\mathcal{D}},\\ \boldsymbol{r}_{\mathcal{S}}\,=\,\tilde{\boldsymbol{r}}_{\mathcal{S}}}} \max_{\substack{\boldsymbol{r}'\,\notin\,\mathcal{R}_{\mathcal{D}},\\ \boldsymbol{r}'_{\mathcal{S}}\,=\,\tilde{\boldsymbol{r}}_{\mathcal{S}}}} \exp\{-NE_{i\mathcal{D}}(\mathcal{S},\boldsymbol{r},\boldsymbol{r}')\} \Bigg\},
\end{aligned}
\tag{3.35}
$$

where $E_{m\mathcal{D}}(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}})$ and $E_{i\mathcal{D}}(\mathcal{S},\boldsymbol{r},\boldsymbol{r}')$ are given by,

$$
\begin{aligned}
E_{m\mathcal{D}}(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}}) = {} & \max_{0<\rho\leq 1} -\rho \sum_{k\in\mathcal{D}\setminus\mathcal{S}} \tilde{r}_k + \max_{0<s\leq 1} -\log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k\in\mathcal{S}} P_{X|r_k}(X_k) \\
& \times \left(\sum_{\boldsymbol{X}_{\mathcal{D}\setminus\mathcal{S}}} \prod_{k\in\mathcal{D}\setminus\mathcal{S}} P_{X|r_k}(X_k) P(Y|\boldsymbol{X}_{\mathcal{D}},\boldsymbol{r}_{\bar{\mathcal{D}}})^{1-s} \right) \\
& \times \left(\sum_{\boldsymbol{X}_{\mathcal{D}\setminus\mathcal{S}}} \prod_{k\in\mathcal{D}\setminus\mathcal{S}} P_{X|\tilde{r}_k}(X_k) P(Y|\boldsymbol{X}_{\mathcal{D}},\tilde{\boldsymbol{r}}_{\bar{\mathcal{D}}})^{\frac{s}{\rho}} \right)^{\rho}, \\
E_{i\mathcal{D}}(\mathcal{S},\boldsymbol{r},\boldsymbol{r}') = {} & \max_{0<\rho\leq 1} -\rho \sum_{k\in\mathcal{D}\setminus\mathcal{S}} r_k + \max_{0<s\leq 1-\rho} -\log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k\in\mathcal{S}} P_{X|r_k}(X_k) \\
& \times \left(\sum_{\boldsymbol{X}_{\mathcal{D}\setminus\mathcal{S}}} \prod_{k\in\mathcal{D}\setminus\mathcal{S}} P_{X|r_k}(X_k) P(Y|\boldsymbol{X}_{\mathcal{D}},\boldsymbol{r}_{\bar{\mathcal{D}}})^{\frac{s}{s+\rho}} \right)^{s+\rho} \\
& \times \left(\sum_{\boldsymbol{X}_{\mathcal{D}\setminus\mathcal{S}}} \prod_{k\in\mathcal{D}\setminus\mathcal{S}} P_{X|r'_k}(X_k) P(Y|\boldsymbol{X}_{\mathcal{D}},\boldsymbol{r}'_{\bar{\mathcal{D}}}) \right)^{1-s},
\end{aligned}
\tag{3.36}
$$

with $P(Y|\boldsymbol{X}_{\mathcal{D}},\boldsymbol{r}_{\bar{\mathcal{D}}})$ in the above equations defined as

$$
P(Y|\boldsymbol{X}_{\mathcal{D}},\boldsymbol{r}_{\bar{\mathcal{D}}}) = \sum_{\boldsymbol{X}_{\bar{\mathcal{D}}}} \prod_{k\in\bar{\mathcal{D}}} P_{X|r_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X}).
\tag{3.37}
$$

Proof. Since the decoder regards signals from users not in $\mathcal{D}$ as interference, given that users not in $\mathcal{D}$ choose rate $\boldsymbol{r}_{\bar{\mathcal{D}}}$, the multiple access channel experienced by users in $\mathcal{D}$ is characterized by $P(Y|\boldsymbol{X}_{\mathcal{D}},\boldsymbol{r}_{\bar{\mathcal{D}}})$ as specified in (3.37). The system can therefore be regarded as a random multiple access system with $|\mathcal{D}|$ users communicating over a compound channel characterized by the set $\{P(Y|\boldsymbol{X}_{\mathcal{D}},\boldsymbol{r}_{\bar{\mathcal{D}}})|\forall\boldsymbol{r}_{\bar{\mathcal{D}}}\}$. Consequently, Lemma 3.5.1 is implied directly by Theorem 3.4.1. □

Next, we will come back to the system where the receiver is only interested in the message of user k. We assume that for each user subset $\mathcal{D} \subseteq \{1, \cdots, K\}$ with $k \in \mathcal{D}$, the receiver assigns an operation region $\mathcal{R}_{\mathcal{D}} \subseteq \mathcal{R}$ for the $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder. That is, if the transmission rate $\boldsymbol{r}$ satisfies $\boldsymbol{r} \in \mathcal{R}_{\mathcal{D}}$, the receiver intends to use the $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder to recover the message of user k. It is easy to see that we should have,

$$\mathcal{R} = \bigcup_{\mathcal{D}:\mathcal{D}\subseteq\{1,\cdots,K\},k\in\mathcal{D}} \mathcal{R}_{\mathcal{D}}. \tag{3.38}$$

Assume that the receiver (single-user decoder) carries out all the $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoding operations. The receiver outputs an estimated message $\hat{w}_k$ for user k if at least one $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoder outputs an estimated message, and all estimation outputs of the $(\mathcal{D}, \mathcal{R}_{\mathcal{D}})$-decoders for user k are identical. Otherwise, the receiver reports a collision for user k.

Let the transmitted rate vector be $\boldsymbol{r}$, and the transmitted message of user k be w_k. We define the decoding error probability $P_e(w_k, \boldsymbol{r})$, the collision miss detection probability $\bar{P}_c(w_k, \boldsymbol{r})$ and the system error probability P_{es} as follows,

$$\begin{aligned}
&P_e(w_k, \boldsymbol{r}) = Pr\left\{(\hat{w}_k, \hat{r}_k) \neq (w_k, r_k) | (w_k, \boldsymbol{r})\right\}, \forall (w_k, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}, \\
&\bar{P}_c(w_k, \boldsymbol{r}) = 1 - Pr\left\{\text{“collision”} | (w_k, \boldsymbol{r})\right\} \\
&\qquad\qquad - Pr\left\{(\hat{w}_k, \hat{r}_k) = (w_k, r_k) | (w_k, \boldsymbol{r})\right\}, \forall (w_k, \boldsymbol{r}), \boldsymbol{r} \notin \mathcal{R}, \\
&P_{es} = \max\left\{\max_{(w_k,\boldsymbol{r}),\boldsymbol{r}\in\mathcal{R}} P_e(w_k, \boldsymbol{r}), \max_{(w_k,\boldsymbol{r}),\boldsymbol{r}\notin\mathcal{R}} \bar{P}_c(w_k, \boldsymbol{r})\right\}.
\end{aligned} \tag{3.39}$$

The following theorem gives an upper bound on the achievable system error probability of the single-user decoder.

Theorem 3.5.2. *Consider a K-user random multiple access system over a discrete-time memoryless channel $P_{Y|\boldsymbol{X}}$, with the receiver only interested in recovering the message for user k. Assume the receiver chooses an operation region $\mathcal{R}$. Let σ be an arbitrary partitioning of the operation region $\mathcal{R}$ satisfying*

$$\begin{aligned}
&\mathcal{R} = \bigcup_{\mathcal{D}:\mathcal{D}\subseteq\{1,\cdots,K\},k\in\mathcal{D}} \mathcal{R}_{\mathcal{D}}, \\
&\mathcal{R}_{\mathcal{D}'} \cap \mathcal{R}_{\mathcal{D}} = \phi, \forall \mathcal{D}, \mathcal{D}' \subseteq \{1, \cdots, K\}, \mathcal{D}' \neq \mathcal{D}, k \in \mathcal{D}, \mathcal{D}'.
\end{aligned} \tag{3.40}$$

System error probability of the single-user decoder is upper-bounded by,

$$P_{es} \le \min_{\sigma} \sum_{\mathcal{D}:\mathcal{D}\subseteq\{1,\cdots,K\}, k\in\mathcal{D}} P_{es}(\mathcal{D},\mathcal{R}_{\mathcal{D}}), \tag{3.41}$$

where $P_{es}(\mathcal{D},\mathcal{R}_{\mathcal{D}})$ is the system error probability bound of the $(\mathcal{D},\mathcal{R}_{\mathcal{D}})$-decoder, and can be further bounded by (3.35).

Proof. Because a $(\mathcal{D},\mathcal{R}_{\mathcal{D}})$-decoder can always choose to report a collision even if it can decode the messages, its system error probability can be improved by shrinking the operation region $\mathcal{R}_{\mathcal{D}}$. This implies that the receiver of the random access system should partition its operation region $\mathcal{R}$ into $\mathcal{R}_{\mathcal{D}}$ regions that do not overlap with each other. In other words, replacing (3.38) by (3.40) will not hurt the system error performance. The rest of the proof is implied by Lemma 3.5.1. □

Note that the system error probability bound provided in Theorem 3.5.2 is implicit since the optimal partitioning scheme σ that maximize the right hand side of (3.41) is not specified. To find the optimal partitioning, one essentially needs to compute every single term on the right hand side of (3.41) and (3.35) for all rate options and all user subsets. Because both $E_{m\mathcal{D}}(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}})$ and $E_{i\mathcal{D}}(\mathcal{S},\boldsymbol{r},\boldsymbol{r}')$ defined in (3.36) involve the combinations of two user subsets and two rate vectors, the computational complexity of finding the optimal partitioning scheme is $O\left((2M)^{2K}\right)$.

3.6 Proofs

In this section, the proofs of the main theorems are provided.

3.6.1 Proof of Theorem 3.1.1

To derive the system error probability upper bound, we assume that the receiver uses the decoding algorithm whose decoding criteria are specified in (3.8).

We next define three probability terms that will be extensively used in the probability bound derivation.

First, assume that (w,r) is the transmitted message and rate pair with $r \in \mathcal{R}$. We define $P_{m[r,\tilde{r}]}$ as the probability that the receiver finds another codeword with rate $\tilde{r} \in \mathcal{R}$ that has a likelihood value no worse than the transmitted codeword.

$$P_{m[r,\tilde{r}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)}) \leq P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})\right\}, \qquad (\tilde{w},\tilde{r}) \neq (w,r), \tilde{r} \in \mathcal{R}. \tag{3.42}$$

Second, assume that (w,r) is the transmitted message and rate pair with $r \in \mathcal{R}$. We define P_{tr} as the probability that the likelihood of the transmitted codeword is below a predetermined threshold.

$$P_{tr} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)}) \leq e^{-N\tau_r(\boldsymbol{y})}\right\}, \tag{3.43}$$

where $\tau_r(\boldsymbol{y})$ is a threshold, as a function of r and $\boldsymbol{y}$, that will be optimized later[8].

Third, assume that $(\tilde{w},\tilde{r})$ is the transmitted message and rate pair with $\tilde{r} \notin \mathcal{R}$. We define $P_{i[\tilde{r},r]}$ as the probability that the receiver finds another codeword with rate $r \in \mathcal{R}$ that has a likelihood value above the required threshold.

$$P_{i[\tilde{r},r]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)}) > e^{-N\tau_r(\boldsymbol{y})}\right\}, \qquad (w,r) \neq (\tilde{w},\tilde{r}), r \in \mathcal{R}. \tag{3.44}$$

With these probability definitions, we can upper bound the system error probability $P_{es}^{(N)}$ by

$$P_{es}^{(N)} \leq \max\left\{\max_{r\in\mathcal{R}} \sum_{\tilde{r}\in\mathcal{R}} P_{m[r,\tilde{r}]} + P_{tr}, \max_{\tilde{r}\notin\mathcal{R}} \sum_{r\in R} P_{i[\tilde{r},r]}\right\}. \tag{3.45}$$

Next, we will upper bound each of the probability terms on the right hand side of (3.45).

Step 1: Upper-bounding $P_{m[r,\tilde{r}]}$.

Assume that (w,r) is the transmitted message and rate pair with $r \in \mathcal{R}$. Given $r, \tilde{r} \in \mathcal{R}$, $P_{m[r,\tilde{r}]}$ can be written as

$$P_{m[r,\tilde{r}]} = E_\theta\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})\phi_{m[r,\tilde{r}]}(\boldsymbol{y})\right], \tag{3.46}$$

where $\phi_{m[r,\tilde{r}]}(\boldsymbol{y}) = 1$ if $P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)}) \leq P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})$ for some $(\tilde{w},\tilde{r}) \neq (w,r)$, and $\phi_{m[r,\tilde{r}]}(\boldsymbol{y}) = 0$ otherwise.

[8] Note that the subscript r of $\tau_r(\boldsymbol{y})$ represents the corresponding estimated rate of the receiver output. Although with an abuse of the notation, we occasionally use the same symbol r to denote both the transmitted rate and the corresponding rate estimation at the receiver, it is important to note that we do not assume the receiver should know the transmitted rate.

Revised from Gallager's approach [4], for any $\rho > 0$ and $s > 0$, we can bound $\phi_{m[r,\tilde{r}]}(\boldsymbol{y})$ by

$$\phi_{m[r,\tilde{r}]}(\boldsymbol{y}) \leq \left[\frac{\sum_{\tilde{w},(\tilde{w},\tilde{r})\neq(w,r)} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})^{\frac{s}{\rho}}}{P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s}{\rho}}}\right]^{\rho}, \quad \rho > 0, s > 0. \tag{3.47}$$

Consequently, $P_{m[r,\tilde{r}]}$ is upper bounded by

$$\begin{aligned} P_{m[r,\tilde{r}]} &\leq E_\theta\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})\left[\frac{\sum_{\tilde{w},(\tilde{w},\tilde{r})\neq(w,r)} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})^{\frac{s}{\rho}}}{P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s}{\rho}}}\right]^{\rho}\right] \\ &= E_\theta\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s}\left[\sum_{\tilde{w},(\tilde{w},\tilde{r})\neq(w,r)} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})^{\frac{s}{\rho}}\right]^{\rho}\right] \\ &= \sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s}\right] E_\theta\left[\left[\sum_{\tilde{w},(\tilde{w},\tilde{r})\neq(w,r)} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})^{\frac{s}{\rho}}\right]^{\rho}\right], \end{aligned} \tag{3.48}$$

where in the last step, we can separate the expectation operations due to independence between $\boldsymbol{x}_{(w,r)}$ and $\boldsymbol{x}_{(\tilde{w},\tilde{r})}$.

Now assume that $0 < \rho \leq 1$. Inequality (3.48) can be further bounded by

$$\begin{aligned} P_{m[r,\tilde{r}]} &\leq \sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s}\right] E_\theta\left[\left[\sum_{\tilde{w}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})^{\frac{s}{\rho}}\right]^{\rho}\right] \\ &\leq e^{N\rho\tilde{r}}\sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s}\right]\left[E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})^{\frac{s}{\rho}}\right]\right]^{\rho} \\ &= e^{N\rho\tilde{r}}\left\{\sum_{Y}\left[\sum_{X} P_{X|r}(X)P_{Y|X}(Y|X)^{1-s}\right]\left[\sum_{X} P_{X|\tilde{r}}(X)P_{Y|X}(Y|X)^{\frac{s}{\rho}}\right]^{\rho}\right\}^{N} \end{aligned} \tag{3.49}$$

Since (3.49) holds for all $0 < \rho \leq 1$, $s > 0$, and it is easy to verify that the bound becomes trivial for $s > 1$, we have

$$P_{m[r,\tilde{r}]} \leq \exp\left\{-NE_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}})\right\}, \tag{3.50}$$

where $E_m(\tilde{r}, P_{X|r}, P_{X|\tilde{r}})$ is given in (3.7).

Step 2: Upper-bounding P_{tr}.

Assume that (w, r) is the transmitted message and rate pair with $r \in \mathcal{R}$. Rewrite P_{tr} as

$$P_{tr} = E_\theta\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})\phi_{tr}(\boldsymbol{y})\right], \tag{3.51}$$

where $\phi_{tr}(\boldsymbol{y}) = 1$ if $P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)}) \leq e^{-N\tau_r(\boldsymbol{y})}$, otherwise $\phi_{tr}(\boldsymbol{y}) = 0$. Note that the value of $\tau_r(\boldsymbol{y})$ will be specified later.

For any $s_1 > 0$, we can bound $\phi_{tr}(\boldsymbol{y})$ as

$$\phi_{tr}(\boldsymbol{y}) \leq \frac{e^{-Ns_1\tau_r(\boldsymbol{y})}}{P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{s_1}}, \quad s_1 > 0. \tag{3.52}$$

This yields

$$\begin{aligned} P_{tr} &\leq E_\theta\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s_1} e^{-Ns_1\tau_r(\boldsymbol{y})}\right] \\ &= \sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s_1}\right] e^{-Ns_1\tau_r(\boldsymbol{y})}. \end{aligned} \tag{3.53}$$

We will come back to this inequality later when we optimize $\tau_r(\ \boldsymbol{y})$.

Step 3: Upper-bounding $P_{i[\tilde{r},r]}$.

Assume that $(\tilde{w},\tilde{r})$ is the transmitted message and rate pair with $\tilde{r} \notin \mathcal{R}$. Given $r \in \mathcal{R}$, we first rewrite $P_{i[\tilde{r},r]}$ as

$$P_{i[\tilde{r},r]} = E_\theta\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})\phi_{i[\tilde{r},r]}(\boldsymbol{y})\right], \tag{3.54}$$

where $\phi_{i[\tilde{r},r]}(\boldsymbol{y}) = 1$ if there exists (w,r) with $r \in \mathcal{R}$ to satisfy $P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)}) > e^{-N\tau_r(\boldsymbol{y})}$, otherwise $\phi_{i[\tilde{r},r]}(\boldsymbol{y}) = 0$.

For any $s_2 > 0$ and $\tilde{\rho} > 0$, we can bound $\phi_{i[\tilde{r},r]}(\boldsymbol{y})$ by

$$\phi_{i[\tilde{r},r]}(\boldsymbol{y}) \leq \left[\frac{\sum_w P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}}{e^{-N\frac{s_2}{\tilde{\rho}}\tau_r(\boldsymbol{y})}}\right]^{\tilde{\rho}}, \quad s_2 > 0, \tilde{\rho} > 0. \tag{3.55}$$

This gives,

$$\begin{aligned} P_{i[\tilde{r},r]} &\leq E_\theta\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})\left[\sum_w P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]^{\tilde{\rho}} e^{Ns_2\tau_r(\boldsymbol{y})}\right] \\ &= \sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})\right] E_\theta\left[\left[\sum_w P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]^{\tilde{\rho}}\right] e^{Ns_2\tau_r(\boldsymbol{y})}. \end{aligned} \tag{3.56}$$

Note that we can separate the expectation operators in the last step due to independence between $\boldsymbol{x}_{(w,r)}$ and $\boldsymbol{x}_{(\tilde{w},\tilde{r})}$.

Assume $0 < \tilde{\rho} \leq 1$. Inequality (3.56) leads to

$$\begin{aligned} P_{i[\tilde{r},r]} &\leq \sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})\right]\left[E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]\right]^{\tilde{\rho}} e^{Ns_2\tau_r(\boldsymbol{y})} e^{N\tilde{\rho}r} \\ &\leq \max_{\tilde{r}\notin\mathcal{R}} \sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})\right]\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_r(\boldsymbol{y})} e^{N\tilde{\rho}r}. \end{aligned} \tag{3.57}$$

Note that the bound obtained in the last step is no longer a function of $\tilde{r}$.

Step 4: Choosing $\tau_r(\boldsymbol{y})$.

In this step, we determine the typicality threshold $\tau_r(\boldsymbol{y})$ that optimizes the bounds in (3.53) and (3.57).

Let us define $\tilde{r}^* \notin \mathcal{R}$ as

$$\tilde{r}^* = \operatorname*{argmax}_{\tilde{r}\notin\mathcal{R}} \sum_{\boldsymbol{y}} E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r})})\right]\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_r(\boldsymbol{y})}e^{N\tilde{\rho}r}. \tag{3.58}$$

Given $r \in \mathcal{R}$, $\boldsymbol{y}$, and the auxiliary variables $s_1 > 0$, $s_2 > 0$, $0 < \tilde{\rho} \leq 1$, we choose $\tau_r(\boldsymbol{y})$ such that the following equality holds,

$$\begin{aligned} & E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s_1}\right]e^{-Ns_1\tau_r(\boldsymbol{y})} \\ & = E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r}^*)})\right]\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_r(\boldsymbol{y})}e^{N\tilde{\rho}r}. \end{aligned} \tag{3.59}$$

This is always possible since the left hand side of (3.59) decreases in $\tau_r(\boldsymbol{y})$ while the right hand side of (3.59) increases in $\tau_r(\boldsymbol{y})$.

Equation (3.59) implies

$$e^{-N\tau_r(\boldsymbol{y})} = \frac{\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r}^*)})\right]\right\}^{\frac{1}{s_1+s_2}}\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{\tilde{\rho}}{s_1+s_2}} e^{N\frac{\tilde{\rho}}{s_1+s_2}r}}{\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s_1}\right]\right\}^{\frac{1}{s_1+s_2}}}. \tag{3.60}$$

Substituting (3.60) into (3.53) yields

$$\begin{aligned} P_{tr} \quad \leq \quad & \sum_{\boldsymbol{y}}\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{1-s_1}\right]\right\}^{\frac{s_2}{s_1+s_2}}\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r}^*)})\right]\right\}^{\frac{s_1}{s_1+s_2}} \\ & \times\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{s_1\tilde{\rho}}{s_1+s_2}} e^{N\frac{s_1\tilde{\rho}}{s_1+s_2}r}. \end{aligned} \tag{3.61}$$

Let $s_2 < \tilde{\rho}$ and $s_1 = 1 - \frac{s_2}{\tilde{\rho}}$. Inequality (3.61) becomes

$$P_{tr} \leq \sum_{\boldsymbol{y}}\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{\tilde{\rho}^2}{\tilde{\rho}-(1-\tilde{\rho})s_2}}\left\{E_\theta\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r}^*)})\right]\right\}^{\frac{\tilde{\rho}-s_2}{\tilde{\rho}-(1-\tilde{\rho})s_2}} e^{N\frac{\tilde{\rho}(\tilde{\rho}-s_2)}{\tilde{\rho}-(1-\tilde{\rho})s_2}r}. \tag{3.62}$$

Now do a variable change with $\rho = \frac{\tilde{\rho}(\tilde{\rho}-s_2)}{\tilde{\rho}-(1-\tilde{\rho})s_2}$ and $s = 1 - \frac{\tilde{\rho}-s_2}{\tilde{\rho}-(1-\tilde{\rho})s_2}$, and note that $s + \rho \leq 1$. Inequality (3.62) becomes

$$\begin{aligned} P_{tr} &\leq \sum_{\boldsymbol{y}} \left\{ E_\theta \left[P(\boldsymbol{y}|\boldsymbol{x}_{(w,r)})^{\frac{s}{s+\rho}} \right] \right\}^{s+\rho} \left\{ E_\theta \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{w},\tilde{r}^*)}) \right] \right\}^{1-s} e^{N\rho r} \\ &\leq \max_{\tilde{r}\notin\mathcal{R}} \left\{ \sum_Y \left[\sum_X P_{X|r}(X) P_{Y|X}(Y|X)^{\frac{s}{s+\rho}} \right]^{s+\rho} \right. \\ &\quad \left. \times \left[\sum_X P_{X|\tilde{r}}(X) P_{Y|X}(Y|X) \right]^{1-s} \right\}^N e^{N\rho r}. \end{aligned} \tag{3.63}$$

Following the same derivation, we can see that $P_{i[\tilde{r},r]}$ is also upper-bounded by the right hand side of (3.63). Because (3.63) holds for all $0 < \rho \leq 1$ and $0 < s \leq 1 - \rho$, we have

$$P_{tr}, P_{i[\tilde{r},r]} \leq \max_{\tilde{r}\notin\mathcal{R}} \exp\{-NE_i(r, P_{X|r}, P_{X|\tilde{r}})\}, \tag{3.64}$$

where $E_i(r, P_{X|r}, P_{X|\tilde{r}})$ is given in (3.7).

Finally, substituting (3.50) and (3.64) into (3.45) gives the desired result.

3.6.2 Proof of Theorem 3.2.1

Due to the involvement of multiple users, notations used in this proof are rather complicated. To make the proof easy to follow, we carefully organize the derivations according to the same structure as the proof of Theorem 3.1.1. Because Theorem 3.1.1 is indeed a simplified single-user version of Theorem 3.2.1, it will help significantly if the reader follows the proof of Theorem 3.2.1 by comparing it, step by step, to the proof of Theorem 3.1.1.

We assume the receiver uses the decoding algorithm whose decoding criteria are specified in (3.18). However, to facilitate the derivation, we first need to make a minor revision to the decoding rules.

Given the received channel symbols $\boldsymbol{y}$, the receiver outputs a message and rate vector pair $(\boldsymbol{w}, \boldsymbol{r})$, with $\boldsymbol{r} \in \mathcal{R}$, if for *all* user subsets $\mathcal{S} \subset \{1, \cdots, K\}$, the following two conditions are met.

$$\begin{aligned}
\text{C1R: } & -\frac{1}{N} \log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}\} < -\frac{1}{N} \log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}\}, \\
& \quad \text{for all } (\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}}) \text{ with } \tilde{\boldsymbol{r}} \in \mathcal{R}, (\tilde{\boldsymbol{w}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}), \\
& \quad \text{and } (\tilde{w}_k, \tilde{r}_k) \neq (w_k, r_k), \forall k \notin \mathcal{S}, \\
\text{C2R: } & -\frac{1}{N} \log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}\} < \tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y}).
\end{aligned} \tag{3.65}$$

Note that in Condition C1R, we added the requirements of $(\tilde{\boldsymbol{w}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}})$ and $(\tilde{w}_k, \tilde{r}_k) \neq (w_k, r_k)$, $\forall k \notin \mathcal{S}$. The union of Conditions C1R over all user subsets $\mathcal{S} \subset \{1, \cdots, K\}$ gives Condition C1 in (3.18). In Condition C2R, we assume that the typicality threshold $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ depends on both $\boldsymbol{r}$ and $\mathcal{S}$. By taking the union over $\mathcal{S} \subset \{1, \cdots, K\}$, Condition C2R in (3.65) implies that the typicality threshold in Condition C2 of (3.18) should be set at $\tau_{\boldsymbol{r}}(\boldsymbol{y}) = \min_{\mathcal{S} \subset \{1, \cdots, K\}} \tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$. In the rest of the proof, we will analyze the probabilities and optimize the thresholds $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ separately for different $\mathcal{S}$.

Given a user subset $\mathcal{S} \subset \{1, \cdots, K\}$, we define the following three probability terms that will be extensively used in the probability bound derivation.

First, assume that $(\boldsymbol{w}, \boldsymbol{r})$ is the transmitted message and rate pair with $\boldsymbol{r} \in \mathcal{R}$. We define $P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}$ as the probability that the receiver finds another message and rate pair $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}})$ with $\tilde{\boldsymbol{r}} \in \mathcal{R}$, $(\tilde{\boldsymbol{w}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}})$, and $(\tilde{w}_k, \tilde{r}_k) \neq (w_k, r_k), \forall k \notin \mathcal{S}$, that has a likelihood value no worse than the transmitted codeword.

$$\begin{aligned}
& P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}) \leq P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})\right\}, \\
& (\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}} \in \mathcal{R}, (\tilde{\boldsymbol{w}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k, \tilde{r}_k) \neq (w_k, r_k), \forall k \notin \mathcal{S}.
\end{aligned} \tag{3.66}$$

Second, assume that $(\boldsymbol{w}, \boldsymbol{r})$ is the transmitted message and rate pair with $\boldsymbol{r} \in \mathcal{R}$. We define $P_{t[\boldsymbol{r},\mathcal{S}]}$ as the probability that the likelihood of the transmitted codeword is no larger than the predetermined threshold $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$.

$$P_{t[\boldsymbol{r},\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}) \leq e^{-N\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}\right\}, \tag{3.67}$$

where the threshold $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ will be optimized later[9].

Third, assume that $(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})$ is the transmitted message and rate pair with $\tilde{\boldsymbol{r}} \notin \mathcal{R}$. We define $P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}$ as the probability that the receiver finds another message and rate pair $(\boldsymbol{w},\boldsymbol{r})$ with $\boldsymbol{r} \in \mathcal{R}$, $(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}})$, and $(w_k,r_k) \neq (\tilde{w}_k,\tilde{r}_k), \forall k \notin \mathcal{S}$, that has a likelihood value above the required threshold $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$.

$$\begin{aligned}
&P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}) > e^{-N\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}\right\}, \\
&(\boldsymbol{w},\boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}, (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}), (w_k,r_k) \neq (\tilde{w}_k,\tilde{r}_k), \forall k \notin \mathcal{S}. \qquad (3.68)
\end{aligned}$$

With these probability definitions, we can upper bound the system error probability $P_{es}^{(N)}$ by

$$P_{es}^{(N)} \leq \max\left\{\begin{array}{l} \max_{\boldsymbol{r}\in\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \left[\sum_{\tilde{\boldsymbol{r}}\in\mathcal{R},\tilde{\boldsymbol{r}}_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]} + P_{t[\boldsymbol{r},\mathcal{S}]}\right], \\ \max_{\tilde{\boldsymbol{r}}\notin\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \sum_{\boldsymbol{r}\in\mathcal{R},\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}} P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]} \end{array}\right\}. \qquad (3.69)$$

Next, we will upper bound each of the probability terms on the right hand side of (3.69).

Step 1: Upper-bounding $P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}$.

Assume that $(\boldsymbol{w},\boldsymbol{r})$ is the transmitted message and rate pair with $\boldsymbol{r} \in \mathcal{R}$. Given $\boldsymbol{r},\tilde{\boldsymbol{r}} \in \mathcal{R}$, $P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}$ can be written as

$$P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]} = E_{\boldsymbol{\theta}}\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})\phi_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}(\boldsymbol{y})\right], \qquad (3.70)$$

where $\phi_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}(\boldsymbol{y}) = 1$ if $P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}) \leq P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})$ for some $(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})$, with $(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}})$, and $(\tilde{w}_k,\tilde{r}_k) \neq (w_k,r_k), \forall k \notin \mathcal{S}$. $\phi_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}(\boldsymbol{y}) = 0$ otherwise.

For any $\rho > 0$ and $s > 0$, we can bound $\phi_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}(\boldsymbol{y})$ by

$$\phi_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}(\boldsymbol{y}) \leq \left[\frac{\sum_{\tilde{\boldsymbol{w}},(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}})=(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}),(\tilde{w}_k,\tilde{r}_k)\neq(w_k,r_k),\forall k\notin\mathcal{S}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}}}{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s}{\rho}}}\right]^{\rho}, \quad \rho > 0, s > 0. \qquad (3.71)$$

[9] As in the single-user case, the subscript $\boldsymbol{r}$ of $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ represents the corresponding estimated rate of the receiver output. Note that we do not assume that the receiver should know the transmitted rate.

Consequently, $P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]}$ is upper bounded by

$$\begin{aligned}
P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]} \quad &\leq \quad E_{\boldsymbol{\theta}}\Bigg[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}) \\
&\qquad \times \Bigg[\frac{\sum_{\tilde{\boldsymbol{w}},(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}})=(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}),(\tilde{w}_k,\tilde{r}_k)\neq(w_k,r_k),\forall k\notin\mathcal{S}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}}}{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s}{\rho}}}\Bigg]^{\rho}\Bigg] \\
&= \quad \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}}\Big[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s} \\
&\qquad \times \Bigg[\sum_{\tilde{\boldsymbol{w}},(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}})=(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}),(\tilde{w}_k,\tilde{r}_k)\neq(w_k,r_k),\forall k\notin\mathcal{S}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}}\Bigg]^{\rho}\Bigg] \\
&= \quad \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\Big[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\Big[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s}\Big] E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \\
&\qquad \times \Bigg[\Bigg[\sum_{\tilde{\boldsymbol{w}},(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}})=(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}),(\tilde{w}_k,\tilde{r}_k)\neq(w_k,r_k),\forall k\notin\mathcal{S}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}}\Bigg]^{\rho}\Bigg]\Bigg],
\end{aligned} \tag{3.72}$$

where in the last step, we can take the expectations operations over users not in $\mathcal{S}$ due to independence between the codewords of $(\boldsymbol{w}_{\bar{\mathcal{S}}},\boldsymbol{r}_{\bar{\mathcal{S}}})$ and $(\tilde{\boldsymbol{w}}_{\bar{\mathcal{S}}},\tilde{\boldsymbol{r}}_{\bar{\mathcal{S}}})$.

Now assume that $0<\rho\leq 1$. Inequality (3.72) can be further bounded by

$$\begin{aligned}
P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]} \quad &\leq \quad \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\Big[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\Big[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s}\Big] E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \\
&\qquad \times \Bigg[\Bigg[\sum_{\tilde{\boldsymbol{w}},(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}})=(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}})} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}}\Bigg]^{\rho}\Bigg]\Bigg] \\
&\leq \quad e^{N\rho\sum_{k\notin\mathcal{S}}\tilde{r}_k}\sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\Big[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\Big[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s}\Big] \\
&\qquad \times \Big[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\Big[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}}\Big]\Big]^{\rho}\Big].
\end{aligned} \tag{3.73}$$

Since (3.73) holds for all $0<\rho\leq 1$, $s>0$, and it is easy to verify that the bound becomes trivial for $s>1$, we have

$$P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{S}]} \leq \exp\left\{-NE_m(\mathcal{S},\tilde{\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}})\right\}, \tag{3.74}$$

where $E_m(\mathcal{S},\tilde{\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}})$ is given in (3.17).

Step 2: Upper-bounding $P_{t[\boldsymbol{r},\mathcal{S}]}$.

Assume that $(\boldsymbol{w}, \boldsymbol{r})$ is the transmitted message and rate pair with $\boldsymbol{r} \in \mathcal{R}$. Rewrite $P_{t[\boldsymbol{r},\mathcal{S}]}$ as

$$P_{t[\boldsymbol{r},\mathcal{S}]} = E_{\boldsymbol{\theta}}\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})\phi_{t[\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y})\right], \tag{3.75}$$

where $\phi_{t[\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y}) = 1$ if $P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}) \leq e^{-N\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}$, otherwise $\phi_{t[\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y}) = 0$. Note that the value of $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ will be specified later.

For any $s_1 > 0$, we can bound $\phi_{t[\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y})$ as

$$\phi_{t[\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y}) \leq \frac{e^{-Ns_1\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}}{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{s_1}}, \quad s_1 > 0. \tag{3.76}$$

This yields

$$\begin{aligned} P_{t[\boldsymbol{r},\mathcal{S}]} &\leq E_{\boldsymbol{\theta}}\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s_1} e^{-Ns_1\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}\right] \\ &= \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s_1}\right] e^{-Ns_1\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}\right]. \end{aligned} \tag{3.77}$$

We will come back to this inequality later when we optimize $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$.

Step 3: Upper-bounding $P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}$.

Assume that $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}})$ is the transmitted message and rate pair with $\tilde{\boldsymbol{r}} \notin \mathcal{R}$. Given $\boldsymbol{r} \in \mathcal{R}$, we first rewrite $P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}$ as

$$P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]} = E_{\boldsymbol{\theta}}\left[\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})})\phi_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y})\right], \tag{3.78}$$

where $\phi_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y}) = 1$ if there exists $(\boldsymbol{w}, \boldsymbol{r})$ with $\boldsymbol{r} \in \mathcal{R}$, $(\boldsymbol{w}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{w}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}})$, and $(w_k, r_k) \neq (\tilde{w}_k, \tilde{r}_k), \forall k \notin \mathcal{S}$ to satisfy $P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}) > e^{-N\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}$. Otherwise $\phi_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y}) = 0$.

For any $s_2 > 0$ and $\tilde{\rho} > 0$, we can bound $\phi_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y})$ by

$$\phi_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}(\boldsymbol{y}) \leq \left[\frac{\sum_{\boldsymbol{w},(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}})=(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}),(w_k,r_k)\neq(\tilde{w}_k,\tilde{r}_k),\forall k\notin\mathcal{S}} P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}}}{e^{-N\frac{s_2}{\tilde{\rho}}\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})}}\right]^{\tilde{\rho}}, \quad s_2 > 0, \tilde{\rho} > 0. \tag{3.79}$$

This gives,

$$\begin{aligned}
P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]} \;\; \leq & \;\; \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}} \left[e^{N s_2 \tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}) \right. \\
& \left. \times \left[\sum_{\boldsymbol{w},(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}})=(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}),(w_k,r_k)\neq(\tilde{w}_k,\tilde{r}_k),\forall k\notin\mathcal{S}} P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right]^{\tilde{\rho}} \right] \\
\leq & \;\; \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}) \right] \right. \\
& \left. \times E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[\left[\sum_{\boldsymbol{w},(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}})=(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}})} P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right]^{\tilde{\rho}} \right] e^{N s_2 \tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} \right]. \qquad (3.80)
\end{aligned}$$

Note that we can separate the expectation operators in the last step due to independence between the codewords of $(\boldsymbol{w}_{\bar{\mathcal{S}}}, \boldsymbol{r}_{\bar{\mathcal{S}}})$ and $(\tilde{\boldsymbol{w}}_{\bar{\mathcal{S}}}, \tilde{\boldsymbol{r}}_{\bar{\mathcal{S}}})$.

Assume that $0 < \tilde{\rho} \leq 1$. Inequality (3.80) leads to

$$\begin{aligned}
P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]} \;\; \leq & \;\; \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}) \right] \left\{ E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right] \right\}^{\tilde{\rho}} \right. \\
& \left. \times e^{N s_2 \tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} e^{N \tilde{\rho} \sum_{k\notin\mathcal{S}} r_k} \right] \\
\leq & \;\; \max_{\boldsymbol{r}'\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w}',\boldsymbol{r}')}) \right] \left\{ E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right] \right\}^{\tilde{\rho}} \right. \\
& \left. \times e^{N s_2 \tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} e^{N \tilde{\rho} \sum_{k\notin\mathcal{S}} r_k} \right]. \qquad (3.81)
\end{aligned}$$

Note that the bound obtained in the last step is no longer a function of $\tilde{\boldsymbol{r}}_{\bar{\mathcal{S}}}$.

Step 4: Choosing $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$.

In this step, we determine the typicality threshold $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ that optimizes the bounds in (3.77) and (3.81).

Define $\tilde{\boldsymbol{r}}^* \notin \mathcal{R}$ as

$$\begin{aligned}
\tilde{\boldsymbol{r}}^* \;\; = & \;\; \operatorname*{argmax}_{\boldsymbol{r}'\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w}',\boldsymbol{r}')}) \right] \left\{ E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right] \right\}^{\tilde{\rho}} \right. \\
& \left. \times e^{N s_2 \tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} e^{N \tilde{\rho} \sum_{k\notin\mathcal{S}} r_k} \right]. \qquad (3.82)
\end{aligned}$$

Given $\boldsymbol{r} \in \mathcal{R}$, $\boldsymbol{y}$, and the auxiliary variables $s_1 > 0$, $s_2 > 0$, $0 < \tilde{\rho} \leq 1$, we choose $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ such that the following equality holds.

$$\begin{aligned}
& E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s_1}\right] e^{-Ns_1\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} \\
& = E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}}^*,\tilde{\boldsymbol{r}}^*)})\right]\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}} r_k}. \qquad (3.83)
\end{aligned}$$

This is always possible since the left hand side of (3.83) decreases in $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$ while the right hand side of (3.83) increases in $\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})$.

Equation (3.83) implies

$$\begin{aligned}
e^{-N\tau_{(\boldsymbol{r},\mathcal{S})}(\boldsymbol{y})} = & \frac{\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}}^*,\tilde{\boldsymbol{r}}^*)})\right]\right\}^{\frac{1}{s_1+s_2}}}{\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s_1}\right]\right\}^{\frac{1}{s_1+s_2}}} \\
& \times\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{\tilde{\rho}}{s_1+s_2}} e^{N\frac{\tilde{\rho}}{s_1+s_2}\sum_{k\notin\mathcal{S}} r_k}. \qquad (3.84)
\end{aligned}$$

Substitute (3.84) into (3.77), we get

$$\begin{aligned}
P_{t[\boldsymbol{r},\mathcal{S}]} \leq & \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{1-s_1}\right]\right\}^{\frac{s_2}{s_1+s_2}}\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}}^*,\tilde{\boldsymbol{r}}^*)})\right]\right\}^{\frac{s_1}{s_1+s_2}}\right. \\
& \left.\times\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{s_1\tilde{\rho}}{s_1+s_2}} e^{N\frac{s_1\tilde{\rho}}{s_1+s_2}\sum_{k\notin\mathcal{S}} r_k}\right]. \qquad (3.85)
\end{aligned}$$

Assume that $s_2 < \tilde{\rho}$. Let $s_1 = 1 - \frac{s_2}{\tilde{\rho}}$. Inequality (3.85) becomes

$$\begin{aligned}
P_{t[\boldsymbol{r},\mathcal{S}]} \leq & \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{\tilde{\rho}^2}{\tilde{\rho}-(1-\tilde{\rho})s_2}}\right. \\
& \left.\times\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}}^*)})\right]\right\}^{\frac{\tilde{\rho}-s_2}{\tilde{\rho}-(1-\tilde{\rho})s_2}} e^{N\frac{\tilde{\rho}(\tilde{\rho}-s_2)}{\tilde{\rho}-(1-\tilde{\rho})s_2}\sum_{k\notin\mathcal{S}} r_k}\right]. \qquad (3.86)
\end{aligned}$$

Now do a variable change with $\rho = \frac{\tilde{\rho}(\tilde{\rho}-s_2)}{\tilde{\rho}-(1-\tilde{\rho})s_2}$ and $s = 1 - \frac{\tilde{\rho}-s_2}{\tilde{\rho}-(1-\tilde{\rho})s_2}$, and note that $s+\rho \leq 1$. Inequality (3.86) becomes

$$\begin{aligned}
P_{t[\boldsymbol{r},\mathcal{S}]} \leq & \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})})^{\frac{s}{s+\rho}}\right]\right\}^{s+\rho}\right. \\
& \left.\times\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}}^*,\tilde{\boldsymbol{r}}^*)})\right]\right\}^{1-s} e^{N\rho\sum_{k\notin\mathcal{S}} r_k}\right] \\
\leq & \max_{\boldsymbol{r}'\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}}\left\{\sum_{Y}\sum_{\boldsymbol{X}_{\mathcal{S}}}\prod_{k\in\mathcal{S}} P_{X|r_k}(X_k)\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|r_k}(X_k)P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{s+\rho}}\right)^{s+\rho}\right. \\
& \left.\times\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|r'_k}(X_k)P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})\right)^{1-s}\right\}^{N} e^{N\rho\sum_{k\notin\mathcal{S}} r_k}. \qquad (3.87)
\end{aligned}$$

Following the same derivation, we can see that $P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]}$ is also upper-bounded by the right hand side of (3.87). Because (3.87) holds for all $0 < \rho \leq 1$ and $0 < s \leq 1-\rho$, we have

$$P_{t[\boldsymbol{r},\mathcal{S}]}, P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r},\mathcal{S}]} \leq \max_{\boldsymbol{r}' \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp\{-N E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})\}, \tag{3.88}$$

where $E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})$ is given in (3.17).

Finally, substituting (3.74) and (3.88) into (3.69) gives the desired result.

3.6.3 Proof of Theorem 3.3.4

We first present in the following lemma an achievable error probability bound for a given codeword length N.

Lemma 3.6.1. *Consider K-user random multiple access communication over a discrete-time memoryless channel $P_{Y|\boldsymbol{X}}$. Assume generalized random coding $(\boldsymbol{\mathcal{L}}^{(N)}, \boldsymbol{\gamma}^{(N)})$ with a finite codeword length N and $e^{NR_{\max}}$ codewords in each codebook. Let the codewords of user k be partitioned into M_k classes, with the i^{th} codeword class corresponding to the standard rate interval $(r^U_{k,i-1}, r^U_{k,i}]$. Assume that $r^U_{k,0} < 0 \leq r^U_{k,1} \leq r^U_{k,2} \cdots \leq r^U_{k,M_k} = R_{\max}$. We term $\{r^U_{k,1}, r^U_{k,2}, \cdots, r^U_{k,M_k}\}$ the grid rates of user k. For any rate $r_k \in (r^U_{k,i-1}, r^U_{k,i}]$, we define function $U(r_k) = r^U_{k,i}$, which rounds r_k to its grid rate value. Let $\boldsymbol{U}(\boldsymbol{r})$ be the vector version of the $U(r)$ function. Denote $\boldsymbol{r}^U$ as a rate vector whose entries only take grid rate values of the corresponding users. Given an operation region $\mathcal{R}$ strictly contained in an achievable rate region, system error probability is upper-bounded by*

$$\begin{aligned} P_{es} \leq \max \Bigg\{ & \max_{\boldsymbol{r} \in \mathcal{R}} \sum_{\mathcal{S} \subset \{1,\cdots,K\}} \\ & \Bigg[\sum_{\tilde{\boldsymbol{r}}^U, \tilde{\boldsymbol{r}}^U_{\mathcal{S}} = \boldsymbol{U}(\boldsymbol{r}_{\mathcal{S}})} \exp\{-N\tilde{E}_m(\mathcal{S}, \tilde{\boldsymbol{r}}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}, \forall \tilde{\boldsymbol{r}} \in \mathcal{R}, \boldsymbol{U}(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^U, \tilde{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}})\} \\ & + \max_{\boldsymbol{r}' \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp\{-N\tilde{E}_i(\mathcal{S}, \boldsymbol{U}(\boldsymbol{r}), \boldsymbol{P}_{\boldsymbol{X}|\hat{\boldsymbol{r}}, \forall \hat{\boldsymbol{r}} \in \mathcal{R}, \boldsymbol{U}(\hat{\boldsymbol{r}}) = \boldsymbol{U}(\boldsymbol{r}), \hat{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}'_{\mathcal{S}}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})\} \Bigg], \\ & \max_{\tilde{\boldsymbol{r}} \notin \mathcal{R}} \sum_{\mathcal{S} \subset \{1,\cdots,K\}} \sum_{\boldsymbol{r}^U, \boldsymbol{r}^U_{\mathcal{S}} = \boldsymbol{U}(\tilde{\boldsymbol{r}}_{\mathcal{S}})} \\ & \max_{\boldsymbol{r}' \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \tilde{\boldsymbol{r}}_{\mathcal{S}}} \exp\{-N\tilde{E}_i(\mathcal{S}, \boldsymbol{r}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}, \forall \boldsymbol{r} \in \mathcal{R}, \boldsymbol{U}(\boldsymbol{r}) = \boldsymbol{r}^U, \boldsymbol{r}_{\mathcal{S}} = \boldsymbol{r}'_{\mathcal{S}}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})\} \Bigg\}, \end{aligned} \tag{3.89}$$

where exponents $\tilde{E}_m(\mathcal{S}, \tilde{\boldsymbol{r}}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}, \forall \tilde{\boldsymbol{r}} \in \mathcal{R}, \boldsymbol{U}(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^U, \tilde{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}})$ *and* $\tilde{E}_i(\mathcal{S}, \boldsymbol{r}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}, \forall \boldsymbol{r} \in \mathcal{R}, \boldsymbol{U}(\boldsymbol{r}) = \boldsymbol{r}^U, \boldsymbol{r}_{\mathcal{S}} = \boldsymbol{r}'_{\mathcal{S}}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'})$ *are defined by*

$$\begin{aligned}
&\tilde{E}_m(\mathcal{S}, \tilde{\boldsymbol{r}}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}, \forall \tilde{\boldsymbol{r}} \in \mathcal{R}, \boldsymbol{U}(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^U, \tilde{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}}) = \max_{0<\rho\le 1} -\rho \sum_{k \notin \mathcal{S}} \tilde{r}_k^U + \max_{0<s\le 1} \\
&\qquad -\log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k \in \mathcal{S}} P_{X|r_k}(X_k) \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{1-s} \right) \\
&\qquad \times \min_{\tilde{\boldsymbol{r}} \in \mathcal{R}, \boldsymbol{U}(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^U, \tilde{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}},} \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|\tilde{r}_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{\rho}} \right)^{\rho}, \\
&\tilde{E}_i(\mathcal{S}, \boldsymbol{r}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}, \forall \boldsymbol{r} \in \mathcal{R}, \boldsymbol{U}(\boldsymbol{r}) = \boldsymbol{r}^U, \boldsymbol{r}_{\mathcal{S}} = \boldsymbol{r}'_{\mathcal{S}}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}'}) = \max_{0<\rho\le 1} -\rho \sum_{k \notin \mathcal{S}} r_k^U + \max_{0<s\le 1-\rho} \\
&\qquad -\log \sum_{Y} \sum_{\boldsymbol{X}_{\mathcal{S}}} \prod_{k \in \mathcal{S}} P_{X|r_k}(X_k) \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r'_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X}) \right)^{1-s} \\
&\qquad \times \min_{\boldsymbol{r} \in \mathcal{R}, \boldsymbol{U}(\boldsymbol{r}) = \boldsymbol{r}^U, \boldsymbol{r}_{\mathcal{S}} = \boldsymbol{r}'_{\mathcal{S}}} \left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_k}(X_k) P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{s+\rho}} \right)^{s+\rho}.
\end{aligned} \tag{3.90}$$

The proof of Lemma 3.6.1 is given in Section 3.6.4.

We will now prove Theorem 3.3.4 based on Lemma 3.6.1. Let the sequence of generalized random coding schemes $\{(\boldsymbol{\mathcal{L}}^{(N)}, \boldsymbol{\gamma}^{(N)})\}$ follow asymptotic input distribution $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$. Given a finite codeword length N, the input distribution of $(\boldsymbol{\mathcal{L}}^{(N)}, \boldsymbol{\gamma}^{(N)})$ is denoted by $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{W}^{(N)}}$. We assume that convergence on the sequence of input distributions $\{\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{W}^{(N)}}\}$ to its asymptotic limit $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$ is uniform[10].

Assume that for each user, say user k, we partition its codewords into M_k classes, as described in Lemma 3.6.1. The i^{th} codeword class corresponding to standard rate interval $(r_{k,i-1}^U, r_{k,i}^U]$. Assume that $r_{k,0}^U < 0 \le r_{k,1}^U \le r_{k,2}^U \cdots \le r_{k,M_k}^U = R_{\max}$. For any rate $r_k \in (r_{k,i-1}^U, r_{k,i}^U]$, we define function $U(r_k) = r_{k,i}^U$, which rounds r_k to its grid rate. Let $\boldsymbol{U}(\boldsymbol{r})$ be the vector version of the $U(r_k)$ function. Denote $\boldsymbol{r}^U$ as a rate vector whose entries only take grid rate values of the corresponding users. Given a finite codeword length N, and the operation region $\mathcal{R}$, system error probability is

[10] Note that $\{\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{W}^{(N)}}\}$ is a *deterministic* sequence.

upper-bounded by (3.89) given in Lemma 3.6.1. Let us regard the codebook partitioning as a rate partitioning, specified by $r_{k,0}^U < 0 \leq r_{k,1}^U \leq r_{k,2}^U \cdots \leq r_{k,M_k}^U = R_{\max}$ for user k, $\forall k$. If we fix the rate partitioning and take the codeword length to infinity, we can lower-bound the system error exponent as

$$E_s \geq \min \left\{ \begin{array}{l} \min_{\mathcal{S}\subset\{1,\cdots,K\}} \min_{\boldsymbol{r}\in\mathcal{R},\tilde{\boldsymbol{r}}^U} \tilde{E}_m(\mathcal{S}, \tilde{\boldsymbol{r}}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}},\forall\tilde{\boldsymbol{r}}\in\mathcal{R},\boldsymbol{U}(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^U,\tilde{\boldsymbol{r}}_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}}), \\ \min_{\mathcal{S}\subset\{1,\cdots,K\}} \min_{\tilde{\boldsymbol{r}}\notin\mathcal{R},\boldsymbol{r}^U} \tilde{E}_i(\mathcal{S}, \boldsymbol{r}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r},\forall\boldsymbol{r}\in\mathcal{R},\boldsymbol{U}(\boldsymbol{r})=\boldsymbol{r}^U,\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}}) \end{array} \right\}, \qquad (3.91)$$

where $\tilde{E}_m(\mathcal{S}, \tilde{\boldsymbol{r}}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}},\forall\tilde{\boldsymbol{r}}\in\mathcal{R},\boldsymbol{U}(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^U,\tilde{\boldsymbol{r}}_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}})$ and $\tilde{E}_i(\mathcal{S}, \boldsymbol{r}^U, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r},\forall\boldsymbol{r}\in\mathcal{R},\boldsymbol{U}(\boldsymbol{r})=\boldsymbol{r}^U,\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}}, \boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}}})$ are defined in (3.90).

Define δ as the maximum width of the rate intervals.

$$\delta = \max_{k\in\{1,\cdots,K\}, i\in\{1,\cdots,M_k\}} r_{k,i}^U - r_{k,i-1}^U \qquad (3.92)$$

Because (3.91) holds for any arbitrary rate partitioning, if we first take codeword length N to infinity, and then slowly revise the rate partitioning by taking δ to zero (which means M_k for all k are taking to infinity), and make sure all input distributions within each rate class converge uniformly to a single asymptotic distribution, then (3.91) implies (3.19). Note that the action of "slowly taking δ to zero" is valid since rate partitioning is only used as a tool for error exponent bound derivation. Revision on the rate partitioning does not require any change to the encoding and decoding schemes. The requirement that all input distributions within each rate class should converge uniformly as δ is taken to zero is also valid since the asymptotic input distribution function of each user is only discontinuous at a finite number of rate points.

3.6.4 Proof of Lemma 3.6.1

Since the codewords in each codebook are partitioned into classes, we will prove Lemma 3.6.1 by following steps similar to the proof of Theorem 3.2.1, with revisions on the bounding details due to the fact that input distributions corresponding to codewords within each class can be different. We will not repeat the proof of Theorem 3.2.1, but only explain the necessary revisions. Throughout the proof, whenever we talk about a message and rate pair $(\boldsymbol{W}, \boldsymbol{r})$, we assume $\boldsymbol{r}$ is the standard communication rate of $\boldsymbol{W}$.

We assume a similar decoding algorithm as given in (3.65), with the second condition being revised to

$$\text{C2R: } -\frac{1}{N}\log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})}\} < \tau_{(\boldsymbol{r}_{\mathcal{S}},\boldsymbol{U}(\boldsymbol{r}_{\bar{\mathcal{S}}}))}(\boldsymbol{y}). \tag{3.93}$$

In other words, we assume that the typicality threshold $\tau_{(\boldsymbol{r}_{\mathcal{S}},\boldsymbol{U}(\boldsymbol{r}_{\bar{\mathcal{S}}}))}(\boldsymbol{y})$ is a function of the standard rates for users in $\mathcal{S}$ and a function of the grid rates for users not in $\mathcal{S}$.

Given a user subset $\mathcal{S} \subset \{1, \cdots, K\}$, we define the following three probability terms.

First, assume that $(\boldsymbol{W}, \boldsymbol{r})$ is the transmitted message and rate pair with $\boldsymbol{r} \in \mathcal{R}$. We define $P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}}^U,\mathcal{S}]}$ as the probability that the receiver finds another codeword and rate pair $(\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{r}})$ with $\tilde{\boldsymbol{r}} \in \mathcal{R}$, $\boldsymbol{U}(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^U$, $(\tilde{\boldsymbol{W}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{W}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}})$, and $(\tilde{W}_k, \tilde{r}_k) \neq (W_k, r_k), \forall k \notin \mathcal{S}$, that has a likelihood value no worse than the transmitted codeword. That is

$$\begin{aligned} P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}}^U,\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(W,r)}) \leq P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{W},\tilde{r})})\right\}, (\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}} \in \mathcal{R}, \boldsymbol{U}(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^U, \\ (\tilde{\boldsymbol{W}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{W}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}), (\tilde{W}_k, \tilde{r}_k) \neq (W_k, r_k), \forall k \notin \mathcal{S}. \end{aligned} \tag{3.94}$$

Second, assume that $(\boldsymbol{W}, \boldsymbol{r})$ is the transmitted message and rate pair with $\boldsymbol{r} \in \mathcal{R}$. We define $P_{t[\boldsymbol{r},\mathcal{S}]}$ as in (3.67) except the typicality threshold is replaced by $\tau_{(\boldsymbol{r}_{\mathcal{S}},\boldsymbol{U}(\boldsymbol{r}_{\bar{\mathcal{S}}}))}(\boldsymbol{y})$.

Third, assume that $(\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{r}})$ is the transmitted message and rate pair with $\tilde{\boldsymbol{r}} \notin \mathcal{R}$. We define $P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r}^U,\mathcal{S}]}$ as the probability that the receiver finds another codeword and rate pair $(\boldsymbol{W}, \boldsymbol{r})$ with $\boldsymbol{r} \in \mathcal{R}$, $\boldsymbol{U}(\boldsymbol{r}) = \boldsymbol{r}^U$, $(\boldsymbol{W}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{W}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}})$, and $(W_k, r_k) \neq (\tilde{W}_k, \tilde{r}_k), \forall k \notin \mathcal{S}$, that has a likelihood value above the required threshold $\tau_{(\tilde{\boldsymbol{r}}_{\mathcal{S}},\boldsymbol{r}^U_{\bar{\mathcal{S}}})}(\boldsymbol{y})$. That is

$$\begin{aligned} P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r}^U,\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})}) > e^{-N\tau_{(\tilde{\boldsymbol{r}}_{\mathcal{S}},\boldsymbol{r}^U_{\bar{\mathcal{S}}})}(\boldsymbol{y})}\right\}, (\boldsymbol{W}, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}, \boldsymbol{U}(\boldsymbol{r}) = \boldsymbol{r}^U, \\ (\boldsymbol{W}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{W}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}), (W_k, r_k) \neq (\tilde{W}_k, \tilde{r}_k), \forall k \notin \mathcal{S}. \end{aligned} \tag{3.95}$$

With the probability definitions, we can upper bound the system error probability P_{es} by

$$P_{es} \leq \max\left\{ \begin{array}{l} \max_{\boldsymbol{r}\in\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \left[\sum_{\tilde{\boldsymbol{r}}^U,\tilde{\boldsymbol{r}}^U_{\bar{\mathcal{S}}}=\boldsymbol{U}(\boldsymbol{r}_{\mathcal{S}})} P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}}^U,\mathcal{S}]} + P_{t[\boldsymbol{r},\mathcal{S}]}\right], \\ \max_{\tilde{\boldsymbol{r}}\notin\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \sum_{\boldsymbol{r}^U,\boldsymbol{r}^U_{\bar{\mathcal{S}}}=\boldsymbol{U}(\tilde{\boldsymbol{r}}_{\mathcal{S}})} P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r}^U,\mathcal{S}]} \end{array} \right\}. \tag{3.96}$$

We will then follow similar steps as in the proof of Theorem 3.2.1 to upper bound each of the probability terms on the right hand side of (3.96).

To upper bound $P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}}^U,\mathcal{S}]}$, we assume that $0 < \rho \leq 1$, $0 < s \leq 1$, and get from (3.73) that

$$
\begin{aligned}
P_{m[\boldsymbol{r},\tilde{\boldsymbol{r}}^U,\mathcal{S}]} \leq\ & \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_\mathcal{S}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})})^{1-s} \right] \right. \\
& \times \left. \left[\sum_{\tilde{\boldsymbol{W}},(\tilde{\boldsymbol{W}}_\mathcal{S},\tilde{\boldsymbol{r}}_\mathcal{S})=(\boldsymbol{W}_\mathcal{S},\boldsymbol{r}_\mathcal{S}),\boldsymbol{U}(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^U} E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{W}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}} \right]^{\rho} \right] \right] \\
\leq\ & e^{N\rho \sum_{k\notin\mathcal{S}} \tilde{r}_k^U} \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_\mathcal{S}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})})^{1-s} \right] \right. \\
& \times \left. \left[\max_{\tilde{\boldsymbol{W}},(\tilde{\boldsymbol{W}}_\mathcal{S},\tilde{\boldsymbol{r}}_\mathcal{S})=(\boldsymbol{W}_\mathcal{S},\boldsymbol{r}_\mathcal{S}),\boldsymbol{U}(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^U} E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{W}},\tilde{\boldsymbol{r}})})^{\frac{s}{\rho}} \right] \right]^{\rho} \right] \\
\leq\ & \exp\{-N\tilde{E}_m(\mathcal{S},\tilde{\boldsymbol{r}}^U,\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}},\forall\tilde{\boldsymbol{r}}\in\mathcal{R},\boldsymbol{U}(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^U,\tilde{\boldsymbol{r}}_\mathcal{S}=\boldsymbol{r}_\mathcal{S}})\}, \qquad (3.97)
\end{aligned}
$$

where $\tilde{E}_m(\mathcal{S},\tilde{\boldsymbol{r}}^U,\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}},\boldsymbol{P}_{\boldsymbol{X}|\tilde{\boldsymbol{r}},\forall\tilde{\boldsymbol{r}}\in\mathcal{R},\boldsymbol{U}(\tilde{\boldsymbol{r}})=\tilde{\boldsymbol{r}}^U,\tilde{\boldsymbol{r}}_\mathcal{S}=\boldsymbol{r}_\mathcal{S}})$ is defined in (3.90).

To upper bound $P_{t[\boldsymbol{r},\mathcal{S}]}$, we get from (3.77) for $s_1 > 0$ that

$$
P_{t[\boldsymbol{r},\mathcal{S}]} \leq \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_\mathcal{S}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})})^{1-s_1} \right] e^{-Ns_1\tau_{(\boldsymbol{r}_\mathcal{S},\boldsymbol{U}(\boldsymbol{r}_{\bar{\mathcal{S}}}))}(\boldsymbol{y})} \right]. \qquad (3.98)
$$

To upper bound $P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r}^U,\mathcal{S}]}$, we get from (3.81) for $s_2 > 0$ and $0 < \tilde{\rho} \leq 1$ that

$$
\begin{aligned}
P_{i[\tilde{\boldsymbol{r}},\boldsymbol{r}^U,\mathcal{S}]} \leq\ & \sum_{\boldsymbol{y}} e^{Ns_2\tau_{(\tilde{\boldsymbol{r}}_\mathcal{S},\boldsymbol{r}^U_{\bar{\mathcal{S}}})}(\boldsymbol{y})} E_{\boldsymbol{\theta}_\mathcal{S}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{W}},\tilde{\boldsymbol{r}})}) \right] \right. \\
& \times \left. \left\{ \sum_{(\boldsymbol{W},\boldsymbol{r}),\boldsymbol{r}_\mathcal{S}=\tilde{\boldsymbol{r}}_\mathcal{S},\boldsymbol{U}(\boldsymbol{r}_{\bar{\mathcal{S}}})=\boldsymbol{r}^U_{\bar{\mathcal{S}}}} E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right] \right\}^{\tilde{\rho}} \right] \\
\leq\ & \sum_{\boldsymbol{y}} e^{Ns_2\tau_{(\tilde{\boldsymbol{r}}_\mathcal{S},\boldsymbol{r}^U_{\bar{\mathcal{S}}})}(\boldsymbol{y})} e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}} r_k^U} E_{\boldsymbol{\theta}_\mathcal{S}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{W}},\tilde{\boldsymbol{r}})}) \right] \right. \\
& \times \left. \left\{ \max_{(\boldsymbol{W},\boldsymbol{r}),\boldsymbol{r}_\mathcal{S}=\tilde{\boldsymbol{r}}_\mathcal{S},\boldsymbol{U}(\boldsymbol{r}_{\bar{\mathcal{S}}})=\boldsymbol{r}^U_{\bar{\mathcal{S}}}} E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right] \right\}^{\tilde{\rho}} \right] \\
\leq\ & \max_{\boldsymbol{r}'\notin\mathcal{R},\boldsymbol{r}'_\mathcal{S}=\tilde{\boldsymbol{r}}_\mathcal{S}} \sum_{\boldsymbol{y}} e^{Ns_2\tau_{(\tilde{\boldsymbol{r}}_\mathcal{S},\boldsymbol{r}^U_{\bar{\mathcal{S}}})}(\boldsymbol{y})} e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}} r_k^U} E_{\boldsymbol{\theta}_\mathcal{S}} \left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W}',\boldsymbol{r}')}) \right] \right. \\
& \times \left. \left\{ \max_{(\boldsymbol{W},\boldsymbol{r}),\boldsymbol{r}_\mathcal{S}=\tilde{\boldsymbol{r}}_\mathcal{S},\boldsymbol{U}(\boldsymbol{r}_{\bar{\mathcal{S}}})=\boldsymbol{r}^U_{\bar{\mathcal{S}}}} E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}} \left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{W},\boldsymbol{r})})^{\frac{s_2}{\tilde{\rho}}} \right] \right\}^{\tilde{\rho}} \right]. \qquad (3.99)
\end{aligned}
$$

Next, by following a derivation similar to Step 4 in the proof of Theorem 3.2.1, we can optimize (3.98) and (3.99) jointly over $\tau_{(\tilde{\boldsymbol{r}}_\mathcal{S},\boldsymbol{r}^U_{\bar{\mathcal{S}}})}(\boldsymbol{y})$ to obtain the desired result.

3.6.5 Proof of Theorem 3.4.1

We assume that the following decoding algorithm is used at the receiver. Given the received channel output symbols $\boldsymbol{y}$, the receiver outputs a message and rate vector pair $(\boldsymbol{w}, \boldsymbol{r})$ together with a channel realization $P_{Y|\boldsymbol{X}}$ such that $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}$ if the following condition is satisfied, for all user subsets $\mathcal{S} \subset \{1, \cdots, K\}$,

$$\begin{aligned}
&-\frac{1}{N}\log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{Y|\boldsymbol{X}}\} < -\frac{1}{N}\log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}, \tilde{P}_{Y|\boldsymbol{X}}\},\\
&\quad \text{for all } (\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}), (\tilde{\boldsymbol{w}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k, \tilde{r}_k) \neq (w_k, r_k), \forall k \notin \mathcal{S},\\
&\quad \text{and } (\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}), (\boldsymbol{w}, \boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}_{(\mathcal{S},\boldsymbol{y})}, \text{ with}\\
&\quad \mathcal{R}_{(\mathcal{S},\boldsymbol{y})} = \Big\{ (\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}) \Big| (\tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}) \in \mathcal{R},\\
&\qquad\qquad -\frac{1}{N}\log Pr\{\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}, \tilde{P}_{Y|\boldsymbol{X}}\} < \tau_{(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y}) \Big\},
\end{aligned} \tag{3.100}$$

where $\tau_{(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S})}(\cdot)$ is a per-determined typicality threshold function of the channel output symbols $\boldsymbol{y}$, associated with the rate and channel realization pair $(\tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}})$ and the user subset $\mathcal{S}$. If there is no codeword satisfying (3.100), the receiver reports a collision. In other words, for a given $\mathcal{S}$, the receiver searches for the subset of codewords with likelihood values larger than the corresponding typicality threshold. If the subset is not empty, the receiver outputs the codeword with the maximum likelihood value as the estimate for this given $\mathcal{S}$. If the estimates for all $\mathcal{S} \subset \{1, \cdots, K\}$ agree with each other, the receiver regards this estimate as the decoding decision and outputs the corresponding decoded message and rate pair. Otherwise, the receiver reports a collision. Note that in (3.100), for given $\mathcal{S}$ and $(\boldsymbol{w}, \boldsymbol{r})$, we only compare the likelihood value of codeword vector $\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}$ with those of the codeword vectors satisfying $(\tilde{\boldsymbol{w}}_{\mathcal{S}}, \tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}}, \boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k, \tilde{r}_k) \neq (w_k, r_k), \forall k \notin \mathcal{S}$. We will first analyze the error performance for each user subset $\mathcal{S}$ and then derive the overall error performance by taking the union over all $\mathcal{S}$.

Given a user subset $\mathcal{S} \subset \{1, \cdots, K\}$, we define the following probability terms.

First, assume that $(\boldsymbol{w}, \boldsymbol{r})$ is transmitted over channel $P_{Y|\boldsymbol{X}}$, with $(\boldsymbol{r}, P_{Y|\boldsymbol{X}}) \in \mathcal{R}$. Let $P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]}$ be the probability that the likelihood value of the transmitted codeword vector over the channel $P_{Y|\boldsymbol{X}}$ is no larger than the corresponding typicality threshold,

$$P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{Y|\boldsymbol{X}}) \le e^{-N\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})}\right\}. \tag{3.101}$$

Define $P_{m[(\boldsymbol{r},P_{Y|\boldsymbol{X}}),(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),\mathcal{S}]}$ as the probability that the likelihood value of the transmitted codeword vector over the channel realization $P_{Y|\boldsymbol{X}}$ is no larger than that of another codeword $(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})$ with $(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k,\tilde{r}_k) \ne (w_k,r_k), \forall k \notin \mathcal{S}$, over channel $\tilde{P}_{Y|\boldsymbol{X}}$ with $(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}) \in \mathcal{R}$,

$$\begin{aligned}&P_{m[(\boldsymbol{r},P_{Y|\boldsymbol{X}}),(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{Y|\boldsymbol{X}}) \le P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}, \tilde{P}_{Y|\boldsymbol{X}})\right\}\\ &(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}), (\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}) \in \mathcal{R}, (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k,\tilde{r}_k) \ne (w_k,r_k), \forall k \notin \mathcal{S}.\end{aligned} \tag{3.102}$$

Second, assume that $(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})$ is transmitted over channel $\tilde{P}_{Y|\boldsymbol{X}}$, with $(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}) \notin \mathcal{R}$. Define $P_{i[(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}),\mathcal{S}]}$ as the probability that the decoder finds a codeword $(\boldsymbol{w},\boldsymbol{r})$ with $(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}), (w_k,r_k) \ne (\tilde{w}_k,\tilde{r}_k), \forall k \notin \mathcal{S}$, over channel $P_{Y|\boldsymbol{X}}$ with $(\boldsymbol{r},P_{Y|\boldsymbol{X}}) \in \mathcal{R}$, such that its likelihood value is larger than the corresponding typicality threshold,

$$\begin{aligned}&P_{i[(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}),\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{Y|\boldsymbol{X}}) > e^{-N\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})}\right\},\\ &(\boldsymbol{w},\boldsymbol{r},P_{Y|\boldsymbol{X}}), (\boldsymbol{r},P_{Y|\boldsymbol{X}}) \in \mathcal{R}, (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}), (w_k,r_k) \ne (\tilde{w}_k,\tilde{r}_k), \forall k \notin \mathcal{S}.\end{aligned} \tag{3.103}$$

With the above probability definitions, by applying the union bound over all $\mathcal{S}$, we can upper-bound the system error probability by

$$\begin{aligned}P_{es} \le \max\Bigg\{ &\max_{(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}})\notin\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \sum_{(\boldsymbol{r},P_{Y|\boldsymbol{X}})\in\mathcal{R},\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}} P_{i[(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}),\mathcal{S}]},\\ &\max_{(\boldsymbol{r},P_{Y|\boldsymbol{X}})\in\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \left[P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]} + \sum_{(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}})\in\mathcal{R},\tilde{\boldsymbol{r}}_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} P_{m[(\boldsymbol{r},P_{Y|\boldsymbol{X}}),(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),\mathcal{S}]}\right]\Bigg\}\end{aligned} \tag{3.104}$$

Next, we will derive individual upper-bounds for each of the probability terms on the right hand side of (3.104).

By following a derivation similar to (3.70)-(3.74) in Section 3.6.2 (Proof of Theorem 3.2.1), we can bound $P_{m[(\boldsymbol{r},P_{Y|\boldsymbol{X}}),(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),\mathcal{S}]}$ by,

$$P_{m[(\boldsymbol{r},P_{Y|\boldsymbol{X}}),(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),\mathcal{S}]} \le \exp\left\{-NE_m(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}},P_{Y|\boldsymbol{X}},\tilde{P}_{Y|\boldsymbol{X}})\right\}, \tag{3.105}$$

where $E_m(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}},P_{Y|\boldsymbol{X}},\tilde{P}_{Y|\boldsymbol{X}})$ is given in (3.26).

Similarly, by using the same bounding techniques as in (3.75)-(3.77) and (3.78)-(3.81) in Section 3.6.2 (Proof of Theorem 3.2.1), we can upper bound $P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]}$ by, for $s_1 > 0$,

$$P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]} \le \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{1-s_1}\right]e^{-Ns_1\tau_{(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})}\right], \tag{3.106}$$

and upper bound $P_{i[(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}),\mathcal{S}]}$ by, for $s_2 > 0, 0 < \tilde{\rho} \le 1$

$$\begin{aligned} P_{i[(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}),\mathcal{S}]} \le & \max_{(\boldsymbol{r}',P'_{Y|\boldsymbol{X}})\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w}',\boldsymbol{r}')},P'_{Y|\boldsymbol{X}})\right]\right. \\ & \left.\times E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left\{\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})}e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}}r_k}\right]. \end{aligned} \tag{3.107}$$

The value of $\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})$ can be determined by jointly optimizing the bounds in (3.106) and (3.107). Consequently, given $(\boldsymbol{r},P_{Y|\boldsymbol{X}})\in\mathcal{R}$, $\boldsymbol{y}$ and auxiliary variables $s_1>0$, $s_2>0$, $0<\tilde{\rho}\le 1$, we choose $\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})$ such that the following equality is satisfied,

$$\begin{aligned} & E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{1-s_1}\right]e^{-Ns_1\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})} \\ & = E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}}^*,\tilde{\boldsymbol{r}}^*)},\tilde{P}^*_{Y|\boldsymbol{X}})\right] \\ & \times E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left\{\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})}e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}}r_k}. \end{aligned} \tag{3.108}$$

where $(\tilde{\boldsymbol{r}}^*,\tilde{P}^*_{Y|\boldsymbol{X}})$ is defined as[11]

$$\begin{aligned} (\tilde{\boldsymbol{r}}^*,\tilde{P}^*_{Y|\boldsymbol{X}}) = & \operatorname*{argmax}_{(\boldsymbol{r}',P'_{Y|\boldsymbol{X}})\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w}',\boldsymbol{r}')},P'_{Y|\boldsymbol{X}})\right]\right. \\ & \left.\times E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left\{\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})}e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}}r_k}\right]. \end{aligned} \tag{3.109}$$

Finding a solution for (3.108) is always possible since that the left hand side of (3.108) decreases with $\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})$, while the right hand side of (3.108) increases with $\tau_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})$. This yields

[11] Although the notation of $\tilde{\boldsymbol{w}}^*$ is used in (3.108), the result is actually invariant to any choice of the message vector.

the desired typicality threshold, denoted by $\tau^*_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})$, which gives

$$\begin{aligned} e^{-N\tau^*_{(\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S})}(\boldsymbol{y})} &= \frac{\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}}^*,\tilde{\boldsymbol{r}}^*)},\tilde{P}^*_{Y|\boldsymbol{X}})\right]\right\}^{\frac{1}{s_1+s_2}}}{\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{1-s_1}\right]\right\}^{\frac{1}{s_1+s_2}}} \\ &\quad \times E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left\{\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{\tilde{\rho}}{s_1+s_2}} e^{N\frac{\tilde{\rho}}{s_1+s_2}\sum_{k\notin\mathcal{S}} r_k}. \end{aligned} \tag{3.110}$$

Substituting (3.110) into (3.106), we get

$$\begin{aligned} P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]} \leq \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\Bigg[& E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{1-s_1}\right]^{\frac{s_2}{s_1+s_2}} e^{N\frac{s_1\tilde{\rho}}{s_1+s_2}\sum_{k\notin\mathcal{S}} r_k} \\ & \times\left\{E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}}^*,\tilde{\boldsymbol{r}}^*)},\tilde{P}^*_{Y|\boldsymbol{X}})\right]\right\}^{\frac{s_1}{s_1+s_2}} E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left\{\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\frac{s_1\tilde{\rho}}{s_1+s_2}}\Bigg]. \end{aligned} \tag{3.111}$$

Let $s_2 < \tilde{\rho}$ and $s_1 = 1-\frac{s_2}{\tilde{\rho}}$, and then do a variable change with $\rho = \frac{\tilde{\rho}(\tilde{\rho}-s_2)}{\tilde{\rho}-(1-\tilde{\rho})s_2}$ and $s = 1-\frac{\tilde{\rho}-s_2}{\tilde{\rho}-(1-\tilde{\rho})s_2}$. Consequently, inequality (3.111) becomes,

$$\begin{aligned} P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]} \leq & \max_{(\boldsymbol{r}',P'_{Y|\boldsymbol{X}})\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} e^{N\rho\sum_{k\notin\mathcal{S}} r_k} \Bigg\{\sum_{Y}\sum_{\boldsymbol{X}_{\mathcal{S}}}\prod_{k\in\mathcal{S}} P_{X|r_k}(X_k) \\ & \times\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|r_k}(X_k)P_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})^{\frac{s}{s+\rho}}\right)^{s+\rho}\left(\sum_{\boldsymbol{X}_{\bar{\mathcal{S}}}}\prod_{k\notin\mathcal{S}} P_{X|r'_k}(X_k)P'_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})\right)^{1-s}\Bigg\}^N \end{aligned} \tag{3.112}$$

Similarly, we can obtain the same upper bound for $P_{i[(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}),\mathcal{S}]}$ as given at the right hand side of (3.112). Since (3.112) holds for all $0 < \rho \leq 1$ and $0 < s \leq 1-\rho$, we have

$$\begin{aligned} & P_{t[\boldsymbol{r},P_{Y|\boldsymbol{X}},\mathcal{S}]}, P_{i[(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}),(\boldsymbol{r},P_{Y|\boldsymbol{X}}),\mathcal{S}]} \\ & \leq \max_{(\boldsymbol{r}',P'_{Y|\boldsymbol{X}})\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \exp\left\{-NE_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',P_{Y|\boldsymbol{X}},P'_{Y|\boldsymbol{X}})\right\}, \end{aligned} \tag{3.113}$$

where $E_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',P_{Y|\boldsymbol{X}},P'_{Y|\boldsymbol{X}})$ is given in (3.26).

By substituting (3.105) and (3.113) into (3.104), we get the desired result.

3.6.6 Proof of Theorem 3.4.3

We assume that the following decoding algorithm is used at the receiver. Given the channel output sequence $\boldsymbol{y}$, the receiver outputs a message and rate vector pair $(\boldsymbol{w},\boldsymbol{r})$ together with a channel class $\mathcal{F}$ such that $(\boldsymbol{r},\mathcal{F})\in\mathcal{R}$ if for all user subset $\mathcal{S}\subset\{1,\cdots,K\}$, the following condition is satisfied,

$$\begin{aligned}
&-\frac{1}{N}\log Pr\left\{\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P^{\mathcal{F}}_{\min}\right\} < -\frac{1}{N}\log Pr\left\{\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}, P^{\tilde{\mathcal{F}}}_{\max}\right\}, \\
&\quad \text{for all } (\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}), (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k,\tilde{r}_k) \neq (w_k,r_k), \forall k \notin \mathcal{S}, \\
&\quad \text{and } (\boldsymbol{w},\boldsymbol{r},P^{\mathcal{F}}_{\min}), (\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}},P^{\tilde{\mathcal{F}}}_{\max}) \in \mathcal{R}_{(\mathcal{S},\boldsymbol{y})}, \text{ with} \\
&\quad \mathcal{R}_{(\mathcal{S},\boldsymbol{y})} = \Big\{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}},P^{\tilde{\mathcal{F}}})|(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}) \in \mathcal{R}, \\
&\qquad\qquad -\frac{1}{N}Pr\left\{\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}, P^{\tilde{\mathcal{F}}}\right\} < \tau_{(\tilde{\boldsymbol{r}},P^{\tilde{\mathcal{F}}},\mathcal{S})}(\boldsymbol{y})\Big\},
\end{aligned} \tag{3.114}$$

where $\tau_{(\tilde{\boldsymbol{r}},P^{\tilde{\mathcal{F}}},\mathcal{S})}(\cdot)$ is the typicality threshold function. Again, we will first analyze the error performance for each individual $\mathcal{S}$ and then derive the overall error performance by taking the union over all $\mathcal{S}$.

For a given user subset $\mathcal{S} \subset \{1,\cdots,K\}$, the following probability terms are defined.

First, assume that $(\boldsymbol{w},\boldsymbol{r})$ is transmitted over channel $P_{Y|\boldsymbol{X}} \in \mathcal{F}$, with $(\boldsymbol{r},\mathcal{F}) \in \mathcal{R}$. Let $P_{t[\boldsymbol{r},\mathcal{F},P_{Y|\boldsymbol{X}},\mathcal{S}]}$ be the probability that the likelihood value of the transmitted codeword vector calculated using $P^{\mathcal{F}}_{\min}$ is no larger than the corresponding typicality threshold,

$$P_{t[\boldsymbol{r},\mathcal{F},P_{Y|\boldsymbol{X}},\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P^{\mathcal{F}}_{\min}) \leq e^{-N\tau_{(\boldsymbol{r},P^{\mathcal{F}}_{\min},\mathcal{S})}(\boldsymbol{y})}\right\}. \tag{3.115}$$

Define $P_{m[(\boldsymbol{r},\mathcal{F}),(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),P_{Y|\boldsymbol{X}},\mathcal{S}]}$ as the probability that the likelihood value of the transmitted codeword vector calculated using $P^{\mathcal{F}}_{\min}$ is no larger than that of another codeword $(\boldsymbol{w},\boldsymbol{r})$ with $(\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k,\tilde{r}_k) \neq (w_k,r_k), \forall k \notin \mathcal{S}$, calculated using $P^{\tilde{\mathcal{F}}}_{\max}$ with $(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}) \in \mathcal{R}$,

$$\begin{aligned}
&P_{m[(\boldsymbol{r},\mathcal{F}),(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),P_{Y|\boldsymbol{X}},\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P^{\mathcal{F}}_{\min}) \leq P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}, P^{\tilde{\mathcal{F}}}_{\max})\right\} \\
&(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}), (\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}) \in \mathcal{R}, (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}) = (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}), (\tilde{w}_k,\tilde{r}_k) \neq (w_k,r_k), \forall k \notin \mathcal{S}.
\end{aligned} \tag{3.116}$$

Second, assume that $(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})$ is transmitted over channel $\tilde{P}_{Y|\boldsymbol{X}} \in \tilde{\mathcal{F}}$, with $(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}) \notin \mathcal{R}$. Define $P_{i[(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),(\boldsymbol{r},\mathcal{F}),\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S}]}$ as the probability that the decoder finds a codeword $(\boldsymbol{w},\boldsymbol{r})$ with $(\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}), (w_k,r_k) \neq (\tilde{w}_k,\tilde{r}_k), \forall k \notin \mathcal{S}$, over channel class $\mathcal{F}$ with $(\boldsymbol{r},\mathcal{F}) \in \mathcal{R}$, such that its likelihood value calculated using $P_{\min}^{\mathcal{F}}$ is larger than the corresponding typicality threshold,

$$\begin{aligned}
&P_{i[(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),(\boldsymbol{r},\mathcal{F}),\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S}]} = Pr\left\{P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{\min}^{\mathcal{F}}) > e^{-N\tau_{(\boldsymbol{r},P_{\min}^{\mathcal{F}},\mathcal{S})}(\boldsymbol{y})}\right\}, \\
&(\boldsymbol{w},\boldsymbol{r},\mathcal{F}), (\boldsymbol{r},\mathcal{F}) \in \mathcal{R}, (\boldsymbol{w}_{\mathcal{S}},\boldsymbol{r}_{\mathcal{S}}) = (\tilde{\boldsymbol{w}}_{\mathcal{S}},\tilde{\boldsymbol{r}}_{\mathcal{S}}), (w_k,r_k) \neq (\tilde{w}_k,\tilde{r}_k), \forall k \notin \mathcal{S}. \qquad (3.117)
\end{aligned}$$

Consequently, the system error probability P_{es} can be upper-bounded using the above probabilities terms as follows,

$$\begin{aligned}
P_{es} \leq \max\Bigg\{ & \max_{(\boldsymbol{r},P_{Y|\boldsymbol{X}}):P_{Y|\boldsymbol{X}}\in\mathcal{F},(\boldsymbol{r},\mathcal{F})\in\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \Bigg[P_{t[\boldsymbol{r},\mathcal{F},P_{Y|\boldsymbol{X}},\mathcal{S}]} \\
& + \sum_{(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}})\in\mathcal{R},\tilde{\boldsymbol{r}}_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} P_{m[(\boldsymbol{r},\mathcal{F}),(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),P_{Y|\boldsymbol{X}},\mathcal{S}]}\Bigg], \\
& \max_{(\tilde{\boldsymbol{r}},\tilde{P}_{Y|\boldsymbol{X}}):\tilde{P}_{Y|\boldsymbol{X}}\in\mathcal{F},(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}})\notin\mathcal{R}} \sum_{\mathcal{S}\subset\{1,\cdots,K\}} \Bigg[\sum_{(\boldsymbol{r},\mathcal{F})\in\mathcal{R},\boldsymbol{r}_{\mathcal{S}}=\tilde{\boldsymbol{r}}_{\mathcal{S}}} P_{i[(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),(\boldsymbol{r},\mathcal{F}),\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S}]}\Bigg]\Bigg\}. \qquad (3.118)
\end{aligned}$$

Note that we have used the union bound over all user subsets $\mathcal{S}$ to obtain the probability bound in (3.118). Next, we will derive individual bound for each of the probability terms on the right hand side of (3.118).

By using the same bounding techniques as in Section 3.6.2 and Section 3.6.5, we can bound $P_{m[(\boldsymbol{r},\mathcal{F}),(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),P_{Y|\boldsymbol{X}},\mathcal{S}]}$ by,

$$\begin{aligned}
P_{m[(\boldsymbol{r},\mathcal{F}),(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),P_{Y|\boldsymbol{X}},\mathcal{S}]} &\leq \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{Y|\boldsymbol{X}})P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{\min}^{\mathcal{F}})^{-s}\right]\right. \\
&\quad \times \left.\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})}, P_{\max}^{\tilde{\mathcal{F}}})^{\frac{s}{\rho}}\right]\right]^{\rho}\right] e^{N\rho\sum_{k\notin\mathcal{S}}\tilde{r}_k} \\
&\leq \exp\left\{-NE_m(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{F},\tilde{\mathcal{F}})\right\}, \qquad (3.119)
\end{aligned}$$

where $E_m(\mathcal{S},\boldsymbol{r},\tilde{\boldsymbol{r}},\mathcal{F},\tilde{\mathcal{F}})$ is given in (3.31). Note that the second inequality in (3.119) is due to the fact that $P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{Y|\boldsymbol{X}}) \leq P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})}, P_{\max}^{\mathcal{F}})$, and the right hand side of (3.119) is not a function of $P_{Y|\boldsymbol{X}}$.

We can also upper bound $P_{t[\boldsymbol{r},\mathcal{F},P_{Y|\boldsymbol{X}},\mathcal{S}]}$ for any $s_1 > 0$ by,

$$\begin{aligned}
P_{t[\boldsymbol{r},\mathcal{F},P_{Y|\boldsymbol{X}},\mathcal{S}]} &\leq \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{Y|\boldsymbol{X}})P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{\min}^{\mathcal{F}})^{-s_1}\right]\right. \\
&\quad \left.\times e^{-Ns_1\tau_{(\boldsymbol{r},P_{\min}^{\mathcal{F}},\mathcal{S})}(\boldsymbol{y})}\right] \\
&\leq \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{\max}^{\mathcal{F}})P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{\min}^{\mathcal{F}})^{-s_1}\right]\right. \\
&\quad \left.\times e^{-Ns_1\tau_{(\boldsymbol{r},P_{\min}^{\mathcal{F}},\mathcal{S})}(\boldsymbol{y})}\right], \qquad (3.120)
\end{aligned}$$

and upper bound $P_{i[(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),(\boldsymbol{r},\mathcal{F}),\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S}]}$ for any $s_2 > 0, 0 < \tilde{\rho} \leq 1$ by,

$$\begin{aligned}
P_{i[(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),(\boldsymbol{r},\mathcal{F}),\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S}]} &\leq \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\tilde{\boldsymbol{w}},\tilde{\boldsymbol{r}})},\tilde{P}_{Y|\boldsymbol{X}})\right]\right. \\
&\quad \left.\times E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left\{\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{\min}^{\mathcal{F}})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_{(\boldsymbol{r},P_{\min}^{\mathcal{F}},\mathcal{S})}(\boldsymbol{y})}e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}}r_k}\right], \\
&\leq \max_{(\boldsymbol{r}',\mathcal{F}')\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \sum_{\boldsymbol{y}} E_{\boldsymbol{\theta}_{\mathcal{S}}}\left[E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w}',\boldsymbol{r}')},P_{\max}^{\mathcal{F}'})\right]\right. \\
&\quad \left.\times E_{\boldsymbol{\theta}_{\bar{\mathcal{S}}}}\left\{\left[P(\boldsymbol{y}|\boldsymbol{x}_{(\boldsymbol{w},\boldsymbol{r})},P_{\min}^{\mathcal{F}})^{\frac{s_2}{\tilde{\rho}}}\right]\right\}^{\tilde{\rho}} e^{Ns_2\tau_{(\boldsymbol{r},P_{\min}^{\mathcal{F}},\mathcal{S})}(\boldsymbol{y})}e^{N\tilde{\rho}\sum_{k\notin\mathcal{S}}r_k}\right] \qquad (3.121)
\end{aligned}$$

Note that the upper bound given in (3.120) is not a function of $P_{Y|\boldsymbol{X}}$. Similarly, the bound in (3.121) is not a function of $\tilde{P}_{Y|\boldsymbol{X}}$.

Optimization of the typicality threshold $\tau_{(\boldsymbol{r},P_{\min}^{\mathcal{F}},\mathcal{S})}$ can be carried out using the similar technique as introduced in (3.108)-(3.110) in Section 3.6.5. By substituting the optimal $\tau_{(\boldsymbol{r},P_{\min}^{\mathcal{F}},\mathcal{S})}$ into (3.120) and (3.121), we get

$$\begin{aligned}
&P_{t[\boldsymbol{r},\mathcal{F},P_{Y|\boldsymbol{X}},\mathcal{S}]}, P_{i[(\tilde{\boldsymbol{r}},\tilde{\mathcal{F}}),(\boldsymbol{r},\mathcal{F}),\tilde{P}_{Y|\boldsymbol{X}},\mathcal{S}]} \\
&\leq \max_{(\boldsymbol{r}',\mathcal{F}')\notin\mathcal{R},\boldsymbol{r}'_{\mathcal{S}}=\boldsymbol{r}_{\mathcal{S}}} \exp\left\{-NE_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',\mathcal{F},\mathcal{F}')\right\}, \qquad (3.122)
\end{aligned}$$

where $E_i(\mathcal{S},\boldsymbol{r},\boldsymbol{r}',\mathcal{F},\mathcal{F}')$ is given in (3.31).

Combining (3.119), (3.122) and (3.118), we obtain

$$\begin{aligned} P_{es} \leq \max \Bigg\{ & \max_{(\boldsymbol{r}, P_{Y|\boldsymbol{X}}): P_{Y|\boldsymbol{X}} \in \mathcal{F}, (\boldsymbol{r}, \mathcal{F}) \in \mathcal{R}} \sum_{\mathcal{S} \subset \{1, \cdots, K\}} \\ & \Bigg[\max_{(\boldsymbol{r}', \mathcal{F}') \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp \left\{ -N E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{r}', \mathcal{F}, \mathcal{F}') \right\} \\ & + \sum_{(\tilde{\boldsymbol{r}}, \tilde{\mathcal{F}}) \in \mathcal{R}, \tilde{\boldsymbol{r}}_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp \left\{ -N E_m(\mathcal{S}, \boldsymbol{r}, \tilde{\boldsymbol{r}}, \mathcal{F}, \tilde{\mathcal{F}}) \right\} \Bigg], \\ & \max_{(\tilde{\boldsymbol{r}}, \tilde{P}_{Y|\boldsymbol{X}}): \tilde{P}_{Y|\boldsymbol{X}} \in \mathcal{F}, (\tilde{\boldsymbol{r}}, \tilde{\mathcal{F}}) \notin \mathcal{R}} \sum_{\mathcal{S} \subset \{1, \cdots, K\}} \\ & \Bigg[\sum_{(\boldsymbol{r}, \mathcal{F}) \in \mathcal{R}, \boldsymbol{r}_{\mathcal{S}} = \tilde{\boldsymbol{r}}_{\mathcal{S}}} \max_{(\boldsymbol{r}', \mathcal{F}') \notin \mathcal{R}, \boldsymbol{r}'_{\mathcal{S}} = \boldsymbol{r}_{\mathcal{S}}} \exp \left\{ -N E_i(\mathcal{S}, \boldsymbol{r}, \boldsymbol{r}', \mathcal{F}, \mathcal{F}') \right\} \Bigg] \Bigg\}. \end{aligned} \tag{3.123}$$

Since the upper bounds given in (3.119) and (3.122) are not functions of individual channels (but functions of channel classes), the right hand side of (3.123) can be simplified to the right hand side of (3.30).

References

[1] C. Shannon, *A mathematical Theory of Communication*, Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, Jul, Oct. 1948.

[2] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd Edition, Wiley Interscience, 2005.

[3] A. Feinstein, *A New Basic Theorem of Information Theory*, IRE Trans. Inform. Theory, IT-4:2-22, 1954.

[4] R. Gallager, *A Simple Derivation of The Coding Theorem and Some Applications*, IEEE Trans. Inform. Theory, Vol. 11, pp. 3-18, Jan. 1965.

[5] A. Feinstein, *Error Bounds in Noisy Channels Without Memory*, EEE Trans. Inform. Theory, Vol. 1, pp. 13-14, Sep. 1955.

[6] R. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons Inc., 1968.

[7] P. Elias, *Coding for Noisy Channels*, IRE Convention Record, No. 4, pp. 37-46, 1955.

[8] R. Fano, *Transmission of Information*, The MIT Press, Cambridge, Mass., and John Wiley and Sons Inc., New York, 1961.

[9] C. Shannon, R. Gallager, and E. Berlekamp, *Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. I*, Information and Control, Vol. 10, pp. 65-103, 1967.

[10] C. Shannon, R. Gallager, and E. Berlekamp, *Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. II*, Information and Control, Vol. 10, pp. 522-552, 1967.

[11] S. Litsyn, *New Upper Bounds on Error Exponents*, IEEE Trans. Inform. Theory, Vol. 45, pp. 385-398, 1999.

[12] G. Forney, *Concatenated Codes*, The MIT Press, 1966.

[13] E. Blokh and V. Zyablov, *Linear Concatenated Codes*, Nauka, Moscow, 1982 (In Russian).

[14] V. Guruswami and P. Indyk, *Linear-Time Encodable/Decodable Codes With Near-Optimal Rate*, IEEE Trans. Inform. Theory, Vol. 51, pp. 3393-3400, Oct. 2005.

[15] C. Shannon, *Two-way Communication Channels*, Proc. 4th Berkeley Symp. Mathematical Statistics and Probability , J. Neyman, Ed. Berkeley, CA: Univ. Calif. Press, Vol. 1, pp. 611-644, 1961.

[16] H. Liao, *Multiple Access Channels*, Ph.D. Dissertation, Univ. Hawaii, Honolulu, HI, 1972.

[17] R. Ahlswede, *The Capacity Region of A Channel With Two Senders and Two Receivers*, Ann. Probab., Vol. 2, pp. 805-814, Oct. 1974.

[18] A. B. Carleial, *A Case Where Interference Does Not Reduce Capacity*, IEEE Trans. Inform. Theory, Vol. IT-21, pp. 569-570, 1975.

[19] V. R. Cadambe and S. A. Jafar, *Interference Alignment and Degrees of Freedom of the K-User Interference Channel*, IEEE Trans. Inform. Theory, Vol. 54, pp. 3425-3441, 2008

[20] T. Cover, *Broadcast Channels*, IEEE Trans. Inform. Theory, Vol. IT-18, pp. 2-4, 1972.

[21] E. C. Van der Meulen, *A Survey of Multi-Way Channels in Information Theory*, IEEE Trans. Inform. Theory, Vol. IT-23, pp. 1-37, 1977.

[22] T. Cover and A. El Gamal, *An Information Theoretic Proof of Hadamard's Inequality*, IEEE Trans. Inform. Theory, Vol. IT-25, pp. 572-584, 1979.

[23] S. Verdú, *Fifty Years of Shannon Theory*, IEEE Trans. Inform. Theory, Vol. 44, pp. 2057-2078, 1998.

[24] J. Byers, M. Luby and A. Rege, *A Digital Fountain Approach to Reliable Distribution of Bulk Data*, ACM SIGCOMM'98, Vancouver, Canada, Sep. 1998.

[25] J. Luo and A. Ephremides, *A New Channel Coding Approach for Random Access with Bursty Traffic*, submitted to IEEE Trans. on Inform. Theory.

[26] D. Mackay, *Fountain Codes,* IEE Proc. Commun., Vol. 152, pp. 1062-1068, Dec. 2005.

[27] M. Luby, *LT codes*, IEEE FOCS'02, Vancouver, Canada, Nov. 2002.

[28] A. Shokrollahi, *Raptor Codes*, IEEE Trans. Inform. Theory, Vol. 52, pp. 2551-2567, Jun. 2006.

[29] O. Etesami and A. Shokrollahi, *Raptor Codes on Binary Memoryless Symmetric Channels*, IEEE Trans. Inform. Theory, Vol. 52, pp. 2033-2051, May 2006.

[30] S. Shamai, I. Teletar and S. Verdú, *Fountain Capacity*, IEEE Trans. Inform. Theory, Vol. 53, pp. 4327-4376, Nov. 2007.

[31] A. Barg and G. Zémor, *Concatenated Codes: Serial and Parallel*, IEEE Trans. Inform. Theory, Vol. 51, No. 5, pp. 1625-1634, May 2005.

[32] I. S. Reed and G. Solomon, *Polynomial Codes over Certain Finite Fields*, Journal of the Society for Industrial and Applied Mathematics (SIAM) Vol.8, pp. 300C304, 1960.

[33] N. Chen and Z. Yan, *Reduced-Complexity ReedCSolomon Decoders Based on Cyclotomic FFTs*, IEEE Signal Processing Letters, Vol.16, pp.279-282, Apr. 2009.

[34] J. Jeng and T. Truong, *On Decoding of Both Errors and Erasures of a Reed-Solomon Code Using an Inverse-free Berlekamp-Massey Algorithm*, IEEE Trans. Communications, Vol. 47, pp. 1488-1494, Oct. 1999.

[35] A. Barg and G. Zémor, *Multilevel Expander Codes*, IEEE ISIT, Adelaide, Austrilia, Sep. 2005.

[36] J. Justesen, *A Class of Constructive Asymptotically Good Algebraic Codes*, IEEE Trans. Inform. Theory, Vol. IT-18, pp. 652-656, Sep. 1972.

[37] A. Barg, G. Forney, *Random Codes: Minimum Distances and Error Exponents*, IEEE Trans. Inform. Theory, Vol. 48, No. 9, pp. 2568-2573, Sep. 2002

[38] Z. Wang and J. Luo, *Approaching Blokh-Zyablov Error Exponent with Linear-Time Encodable/Decodable Codes*, IEEE Communications Letters, Vol. 13, No. 6, pp. 438-440, June 2009.

[39] D. Romik, “Stirling’s Approximation for n!: The Ultimate Short Proof,” The American Mathematical Monthly, Vol. 107, pp. 556-557, Jun.-Jul. 2000.

[40] R. Gallager, *A Perspective on Multiaccess Channels*, IEEE Trans. Inform. Theory, Vol. 31, pp. 124-142, Mar. 1985

[41] A. Ephremides and B. Hajek, *Information Theory and Communication Networks: An Unconsummated Union,* IEEE Trans. Inform. Theory, Vol. 44, pp. 2416-2434, Oct. 1998.

[42] D. Bertsekas and R. Gallager, *Data Network,* 2nd edition, Prentice Hall, 1992.

[43] N. Abramson, *The Aloha system-Another Alternative for Computer Communications*, Proc. Fall Joint Computer Conf., AFIPS Conf., Vol. 37, 1970.

[44] S. Ghez, S. Verdú, and S. Schwartz, *Stability Properties of Slotted ALOHA with Multipacket Reception Capability,* IEEE Trans. Auto. Contr., Vol. 33, pp. 640-649, Jul. 1988.

[45] J. Luo and A. Ephremides, *On The Throughput, Capacity and Stability Regions of Random Multiple Access,* IEEE Trans. Inform. Theory, Vol. 52, pp. 2593-2607, Jun. 2006.

[46] P. Karn, *MACA-A New Channel Access Method for Packet Radio*, Computer Networking Conf., Vol. 9, pp. 134-140, 1990.

[47] Y. Polyanskiy, H. Vincent Poor, and S. Verdú, *Channel Coding Rate in the Finite Blocklength Regime,* IEEE Trans. Inform. Theory, Vol. 56, pp. 2307-2359, May 2010

[48] A. Lapidoth and P. Narayan, *Reliable Communication under Channel Uncertainty,* IEEE Trans. Inform. Theory, Vol. 44, pp. 2148-177, Oct. 1998.

[49] D. Blackwell, L. Breiman, and A. Thomasian, *The capacity of A Class of Channels,* Ann. Math. Statist., Vol. 30, pp. 1229-241, 1959.

[50] J. Wolfowitz, *Simultaneous Channels,* Arch. Rational Mech. and Anal., Vol. 4, pp. 371-86, 1959.

[51] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, New York: Academic, 1981.

[52] Z. Wang and J. Luo, *Achievable Error Exponent of Channel Coding in Random Access Communication*, ISIT, Jun. 2010.